NASA Technical Memorandum 103835

Three-Dimensional Virtual Acoustic Displays

Elizabeth M. Wenzel

July 1991

National Aeronautics and
Space Administration

NASA Technical Memorandum 103835

Three-Dimensional Virtual Acoustic Displays

Elizabeth M. Wenzel, Ames Research Center, Moffett Field, California

July 1991

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

SUMMARY

The development of an alternative medium for displaying information in complex human-machine interfaces is described. The three-dimensional virtual acoustic display is a means for accurately transfering information to a human operator using the auditory modality; it combines directional and semantic characteristics to form naturalistic representations of dynamic objects and events in remotely-sensed or simulated environments. Although the technology can stand alone, it is envisioned as a component of a larger multisensory environment and will no doubt find its greatest utility in that context. The general philosophy in the design of the display has been that the development of advanced computer interfaces should be driven first by an understanding of human perceptual requirements, and later by technological capabilities or constraints. In expanding on this view, the paper addresses current and potential uses of virtual acoustic displays, characterizes such displays, reviews recent approaches to their implementation and application, describes the research project at NASA Ames in some detail, and finally outlines some critical research issues for the future.

INTRODUCTION

Rather than focus on the "multi" part of multimedia interfaces, this paper will emphasize the justification and development of a particular medium, the three-dimensional virtual acoustic display. Although the technology can stand alone, it is envisioned as a component of a larger multisensory environment and will no doubt find its greatest utility in that context. The general philosophy in the design of the display has been that the development of advanced computer interfaces should be driven first by an understanding of human perceptual requirements, and later by technological capabilities or constraints. In expanding on this view, I will address why virtual acoustic displays are useful, characterize the abilities of such displays, review some recent approaches to their implementation and application, describe the current research at NASA Ames in some detail, and finally outline some critical research issues for the future. Since these goals are rather ambitious, I apologize in advance for neglecting any important work or issues in an area that seems to be rapidly gaining momentum.

WHY VIRTUAL ACOUSTIC DISPLAYS?

The recent burgeoning of computing technology requires that people learn to interpret increasingly complex systems of information and control increasingly complex machines. One approach to this problem has been to develop direct-manipulation, graphical computer interfaces exemplified by the ubiquitous combination of the desktop metaphor and the mouse. Such spatially-organized interfaces can provide familiarity and consistency across applications, thus avoiding much of the task-dependent learning of the older text-oriented displays. Lately, a considerable amount of attention has been devoted to a more ambitious type of reconfigurable interface called the virtual display. Despite the oft-touted "revolutionary" nature of this field, the research has many antecedents in previous

work in three-dimensional computer graphics, interactive input/output devices, and simulation technology. Some of the earliest work in virtual interfaces was done by Sutherland (1968) using binocular head-mounted displays. Sutherland characterized the goal of virtual interface research, stating, "The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real." As technology has advanced, virtual displays have adopted a three-dimensional spatial organization, in order to provide a more natural means of accessing and manipulating information. A few projects have taken the spatial metaphor to its limit by directly involving the operator in a data environment (e.g., Furness, 1986; Brooks, 1988; Fisher et al., 1988). For example, Brooks (1988) and his colleagues have worked on a three-dimensional interface in which a chemist can visually and manually interact with a virtual model of a drug compound, attempting to discover the bonding site of a molecule by literally seeing and feeling the interplay of the chemical forces at work. It seems that the kind of "artificial reality" once relegated solely to the specialized world of the cockpit simulator is now being seen as the next step in interface development for many types of advanced computing applications (Foley, 1987).

Often the only modalities available for interacting with complex information systems have been visual and manual. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel (e.g., Garner, 1949; Deatherage, 1972; Doll et al., 1986). Most recently, attention has been devoted to the use of non-speech audio as an interface medium (Patterson, 1982; Gaver, 1986; Begault and Wenzel, 1990; Blattner et al.,1989; Buxton et al., 1989). For example, auditory signals are detected more quickly than visual signals and tend to produce an alerting or orienting response (Mowbray and Gebhard, 1961; Patterson, 1982). These characteristics are probably responsible for the most prevalent use of non-speech audio in simple warning systems, such as the malfunction alarms used in aircraft cockpits or the siren of an ambulance. Another advantage of audition is that it is primarily a temporal sense and we are extremely sensitive to changes in an acoustic signal over time (Mowbray and Gebhard, 1961; Kubovy, 1981). This feature tends to bring a new acoustical event to our attention and conversely, allows us to relegate sustained or uninformative sounds to the background. Thus audio is particularly suited to monitoring state changes over time, for example, when a car engine suddenly begins to malfunction.

Non-speech signals have the potential to provide an even richer display medium if they are carefully designed with human perceptual abilities in mind. Just as a movie with sound is much more compelling and informationally-rich than a silent film, so could a computer interface be enhanced by an appropriate "sound track" to the task at hand. If used properly, sound need not be distracting or cacophonous or merely uninformative. Principles of design for auditory icons and auditory symbologies can be gleaned from the fields of music (Deutsch, 1982; Blattner et al., 1989), psychoacoustics (Carterette and Friedman, 1978; Patterson, 1982), and psychological studies of the acoustical determinants of perceptual organization (Bregman, 1981; 1990; Kubovy, 1981; Buxton et al., 1989). For example, following from Gibson's (1979) ecological approach to perception, one can conceive of the audible world as a collection of acoustic "objects." Various acoustic features, such as temporal onsets and offsets, timbre, pitch, intensity, and rhythm, can specify the identities of the objects and convey meaning about discrete events or ongoing actions in the world and their relationships to one another. One could systematically manipulate these features, effectively creating an auditory symbology which operates on a continuum from "literal" everyday sounds, such as the clunk of mail in

your mailbox (e.g., Gaver's "Sonic Finder," 1986), to a completely abstract mapping of statistical data into sound parameters (Bly, 1982; Smith et al., 1990; Blattner et al., 1989).

Such a display could be further enhanced by taking advantage of the auditory system's ability to segregate, monitor, and switch attention among simultaneous sources of sound (Mowbray and Gebhard, 1961). One of the most important determinants of acoustic segregation is an object's location in space (Kubovy and Howard, 1976; Bregman, 1981, 1990; Deutsch, 1982).

A three-dimensional auditory display may be most usefully applied in contexts where the representation of spatial information is important, particularly when visual cues are limited or absent and workload is high. Such displays can potentially enhance information transfer by combining directional with iconic information in a quite naturalistic representation of dynamic objects in the interface. Borrowing a term from Gaver (1986), an obvious aspect of "everyday listening" is the fact that we live and listen in a three-dimensional world. A primary advantage of the auditory system is that it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. In fact, I would like to suggest that a good rule of thumb for knowing when to provide acoustic cues is to recall how we naturally use audition to gain information and explore the environment; that is, "the function of the ears is to point the eyes." Thus the auditory system can provide a more coarsely-tuned mechanism to direct the attention of our more finely-tuned visual analyses. For example, Perrott et al. (1991) have recently reported that aurally-guided visual search for a target in a cluttered visual display is superior to unaided visual search, even for objects in the central visual field. Such features will be especially useful in inherently spatial tasks, such as air traffic control (ATC) displays for the tower or cockpit. For example, ATC controllers are being asked to integrate increasingly heavy air traffic into increasingly complex landing patterns, such as the triple parallel approach proposed to maximize the flow of incoming aircraft. Research at NASA Ames, in collaboration with the Federal Aviation Administration, will emphasize two types of acoustic displays because of their conceptual simplicity and the likelihood that they will provide significant benefits to current ATC systems. One example is an ATC display in which the controller hears communications from incoming traffic in positions which correspond to their actual location in the terminal area. In such a display, it should be more immediately obvious to the listener when aircraft are on a potential collision course because they would be heard in their true spatial locations and their routes could be tracked over time. A second example involves alerting systems for ATC. An auditory icon, such as a complex signal with a unique temporal rhythm, could also be used as a warning of urgent situations like potential runway incursions. Again, the signal could be processed to convey true directional information and urgency could be emphasized by placing the warning close to the listener's head, e.g., within the boundaries of their "personal space" (Begault and Wenzel, 1990).

A second advantage of the binaural system, often referred to as the "cocktail party effect", is that it improves the intelligibility of sources in noise and assists in the segregation of multiple sound sources (Cherry, 1953; Bronkhorst and Plomp, 1988). This effect could be critical in applications involving the kind of encoded non-speech messages proposed for scientific "visualization," the acoustic representation of multi-dimensional data (e.g., Bly, 1982; Blattner et al., 1989; Smith et al., 1990), or the development of alternative interfaces for the visually impaired (Edwards, 1989; Loomis et al., 1990). Another aspect of auditory spatial cues is that, in conjunction with the other senses, they can act as potentiators of information in a display. For example, visual and auditory cues together can reinforce the information content of a display and provide a greater sense of presence or

realism in a manner not readily achieved by either modality alone (Colquhoun, 1975; O'Leary and Rhodes, 1984; Warren et al., 1981). Similarly, in direct-manipulation tasks, auditory cues can provide supporting information for the representation of force-feedback (Wenzel et al., 1990), a quite difficult interface problem for multimodal displays which is only beginning to be solved (e.g., Minsky et al., 1990). Intersensory synergism will be particularly useful in telepresence applications, including advanced teleconferencing (Ludwig et al., 1990), shared electronic workspaces (Fisher et al., 1988; Gaver and Smith, 1990), monitoring telerobotic activities in remote or hazardous situations (Wenzel et al., 1990), and entertainment environments (Kendall and Martens, 1984; Kendall and Wilde, 1989; Cooper and Bauck, 1989). Thus, the combination of veridical spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use. Here, the term veridical is used to indicate that spatial cues are both realistic and result in the accurate transfer of information; e.g., the presentation of such cues results in accurate estimates of perceived location by human listeners in psychophysical studies.

From the above considerations, one can attempt to a define a virtual acoustic display and list some of the goals to keep in mind when developing the supporting technology and conducting related perceptual research. A virtual acoustic display is a medium for accurately transfering information to a human operator using the auditory modality; it combines directional and semantic characteristics to form naturalistic representations of dynamic objects and events in remotely-sensed or simulated environments. As with visual displays, this definition does not necessarily mean that the virtual representation must be indistinguishable from reality. Rather, it implies that the display should provide a functional equivalence to human audition in the context of the task to be performed. To achieve this goal, we must know a great deal about our sensory biases; that is, the what, when, and how of the acoustic information used by the human listener. It also means that we must systematically verify that the displays we develop are perceptually viable. Therefore the display must: (1) adequately reproduce the audible spectrum in frequency resolution and dynamic range, (2) present information accurately in three spatial dimensions, (3) be capable of representing multiple sources which can be either static or moving, (4) be real-time and interactive; that is, responsive to the ongoing needs of the user, (5) be head-coupled to provide a stable acoustic environment with dynamic cues appropriately correlated with head motion, and (6) be flexible in the type of acoustic information which can be displayed; for example, real environmental sounds, acoustic icons, speech, or streams of multidimensional auditory patterns or objects. A corollary to this approach is that such a display may potentially be used to enhance normal perceptual capabilities. For example, Durlach (1990; Durlach and Pang, 1986) has proposed that localization cues could be artificially magnified to create a kind of super localization ability.

ANTECEDENTS OF THREE-DIMENSIONAL VIRTUAL ACOUSTIC DISPLAYS

As noted above, the utility of a 3D auditory display greatly depends on the user's ability to localize the various sources of information in auditory space. While compromises obviously have to be made to achieve a practical system, the particular features or limitations of the latest hardware should be considered subservient to human sensory and performance requirements. Thus, designers of such interfaces must carefully consider the acoustic cues needed by listeners for accurate localization and ensure that these cues will be faithfully (or at least adequately, in a human performance sense) trans-

duced by the synthesis device rather than letting current technology drive the implementation. In fact, knowledge about sensory requirements might actually save processing power in some cases and indicate others to which more resources should be devoted.

Psychoacoustical Antecedents

Much of the research on human sound localization is summarized in the classic "duplex theory" which emphasizes the role of two primary cues, interaural differences in time of arrival at low frequencies and interaural differences in intensity at high frequencies (Lord Rayleigh, 1907). However, binaural research over the last 25 years points to serious limitations with this approach (see Blauert, 1983, for an extensive review of spatial hearing). For example, it cannot account for the ability of subjects to localize sounds on the vertical median plane where interaural cues are minimal (Blauert, 1969; Butler and Belendiuk, 1977; Oldfield and Parker, 1986). Similarly, when subjects listen to stimuli over headphones, they are perceived as being inside the head even though interaural temporal and intensity differences appropriate to an external source location are present (Plenge, 1974). Many studies now suggest that deficiencies of the duplex theory reflect the important contribution to localization of the direction-dependent filtering which occurs when incoming sound waves interact with the outer ears or pinnae. Experiments have shown that spectral shaping by the pinnae is highly direction dependent (Shaw, 1974), that the absence of pinna cues degrades localization accuracy (Gardner and Gardner, 1973; Oldfield and Parker, 1984b), and that pinna cues are primarily responsible for externalization or the "outside-the-head" sensation (Plenge, 1974). Such data suggest that perceptually-veridical localization over headphones should be possible if the spectral shaping by the pinnae as well as the interaural difference cues are adequately synthesized.

Approaches to Implementation

Prior to the development of current techniques for synthesizing out-of-head localization, there were some early attempts at creating what we might now call a virtual acoustic display. One of these was the rather amazing pseudophone apparatus (fig. 1) used during World War I for detecting and locating enemy aircraft. It is an early example of the use of enhanced localization cues in the form of large directional pinnae and an expanded interaural axis. A less elaborate display called FLYBAR (FLYing By Auditory Reference) was developed by Forbes (1946) just after World War II. This system used only crude left/right intensity panning along with pitch and temporal pattern changes to indicate turn, bank, and air speed in an acoustic display for instrument flying.

Much later, investigators began to think about simulating veridical auditory localization cues as a way of analyzing and enhancing the listening experience in stereo reproduction, and eventually, to display information. In general, the approaches have concentrated on various means for reproducing the effects of the Head-Related Transfer Function (HRTF); that is, the direction-dependent acoustic effects imposed on an incoming signal by the outer ears. The nature and measurement of the HRTF will be considered later in more detail.

One class of techniques derives from binaural recording and the development of normative manikins, such as the KEMAR (Knowles Electronics, Inc.) and Neumann (e.g., Hudde and Schroter, 1981)

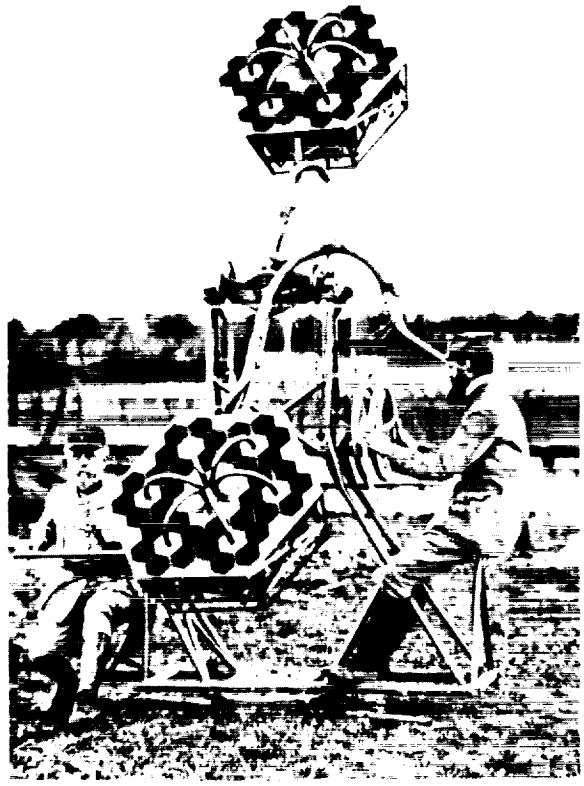


Figure 1. Photo of the pseudophone apparatus used for detecting and localizing aircraft during World War I (from Scientists in Power, Spencer R. Weart, Harvard University Press (Cambridge, Mass.; reproduced with permission, Niels Bohr Library, American Institute of Physics, New York, NY)).

artificial heads, used for applications like assessing concert hall acoustics (see Blauert, 1983). Recent examples of a real time version of this approach in information display include the work by Doll at the Georgia Institute of Technology (Doll et al., 1986) and the Gehring AL100 system developed for the Super Cockpit Project at Wright-Patterson Air Force Base (see Calhoun et al., 1987). These projects used a movable artificial head to simulate moving sources and correlated head-motion. The listener heard headphone signals transduced in the ears of a manikin which was mechanically coupled to that of the listener's own head.

Another type of real time virtual display is the work by Loomis et al. (1990) on a navigation aid for the blind. In this analog system, which worked well in an active tracking task, spatial cues were approximated using various types of simple filters with interaural time and intensity differences dynamically linked to head motion. The display also included simple distance and reverberation cues such as an intensity rolloff with distance and the ratio of direct to reflected energy.

Much of the recent work since the early 80s has been devoted to the measurement and real time digital synthesis of HRTFs. Techniques for creating digital filters based on measurements of finite impulse responses in the ear canals of either individual subjects or artificial heads have been under development since the late 70s. But it is only with the advent of powerful new digital signal-processing (DSP) chips that a few real-time systems have appeared in the last few years in Europe

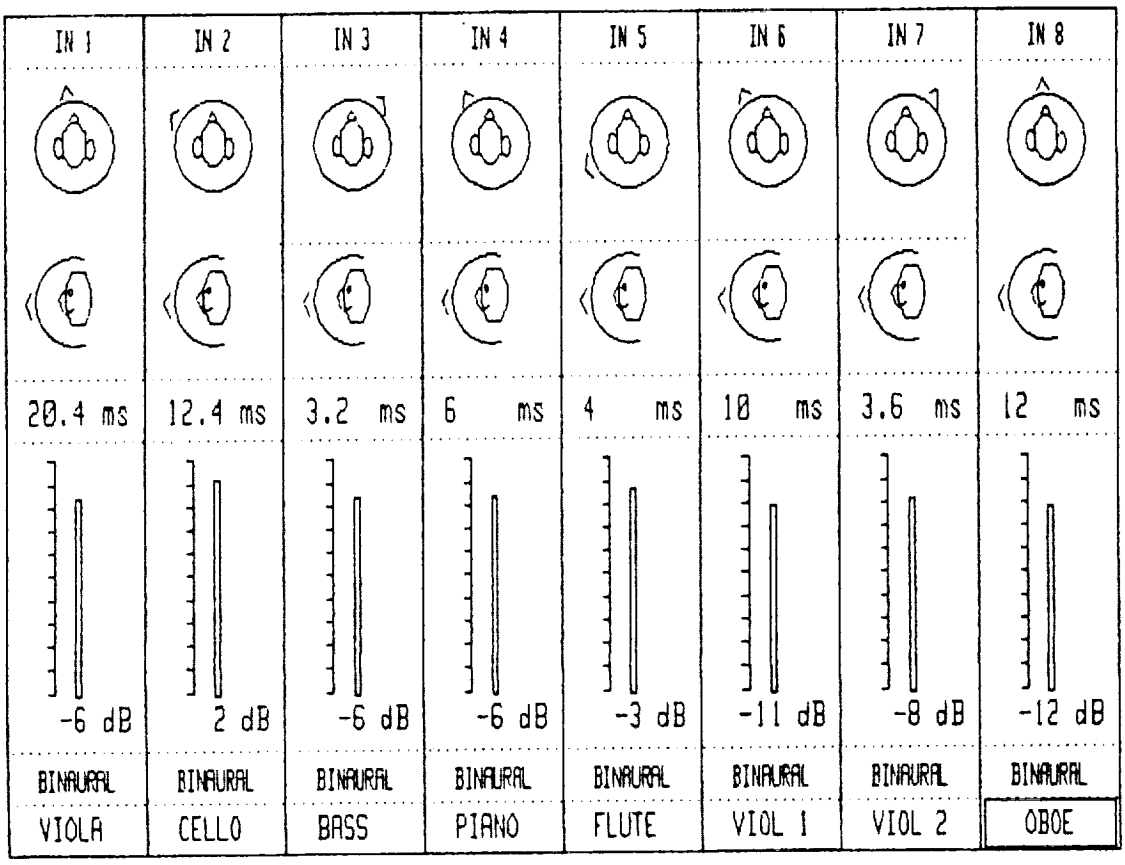
and the United States. In general, these systems are intended for headphone delivery and use time-domain convolution to achieve real time performance.

One example is the Creative Audio Processor, a kind of binaural mixing console, developed by AKG in Austria and based on ideas proposed by Blauert (1984). The CAP 340M is aimed at applications like audio recording, acoustic design, and psychoacoustic research (Persterer, 1989). This particular system is rather large, involving an entire rack of digital signal processors and related hardware. The system is also rather powerful in that up to 32 channels can be independently "spatialized" in azimuth and elevation along with variable simulation of room response characteristics. Figure 2, for example, illustrates the graphical interface of the system for specifying characteristics of the binaural mix for a collection of independently-positioned musical instruments. A collection of HRTFs is offered, derived from measurements taken in the ear canals of both manikins and individual subjects. AKG's original measurements were made by Blauert and his colleagues (Blauert, personal communication). In a new product, which simulates an ideal control room for headphone reproduction, the BAP 1000, the user has the option of having his/her individual transforms programmed onto a PROM card. Interestingly, AKG's literature mentions that best results are achieved with individual transforms. Currently there are plans for the system to be used in an October 1991 mission of the Russian Space Program. The AUDIMIR study examines whether acoustic cues for orientation can eliminate mismatch of auditory and vestibular cues and thus counteract space sickness (AKG Report, Nov. 1989).

Other projects in Europe derive from the efforts of a group of researchers in Germany. This work includes the most recent efforts of Jens Blauert and his colleagues at the Ruhr University at Bochum (Boerger et al., 1977; Lehnert and Blauert, 1989; Posselt et al., 1986). The group at Bochum has been working on a prototype PC-based DSP system, again a kind of binaural mixing console, whose proposed features include real time convolution of HRTFs for up to four sources, interpolation between transforms to simulate motion, and room modeling. The group has devoted quite a bit of effort to measuring HRTFs for both individual subjects and artificial heads (e.g., the Neumann head), as well as developing computer simulations of transforms.

Another researcher in Germany, Klaus Genuit, worked at the Institute of Technology of Aachen and later went on to form his own company, HEAD Acoustics. HEAD Acoustics has also produced a real time, four-channel binaural mixing console and simulator for room acoustics as well as a new version of an artificial head (Gierlich and Genuit, 1989). Genuit's work is particularly notable for his development of a structurally-based model of the acoustic effects of the pinnae (e.g., Genuit, 1986). That is, rather than use individualized HRTFs, Genuit has developed a parameterized, mathematical description (based on Kirchhoff's diffraction integrals) of the acoustic effects of the pinnae, ear canal resonances, torso, shoulder, and head. The effects of the structures have been simplified; for example, the outer ears are modeled as three cylinders of different diameters and length. The parameterization of the model adds some flexibility to this technique and Genuit states that the calculated transforms are within the variability of directly-measured HRTFs.

In the United States, similar projects are currently in progress. For example, at Wright-Patterson Air Force Base, McKinley and Ericson (1988) developed a prototype system which synthesizes a single source in azimuth in real time. The system uses HRTFs based on measurements from a KEMAR manikin made at 1° intervals in azimuth with a head-tracker to achieve source stabilization.



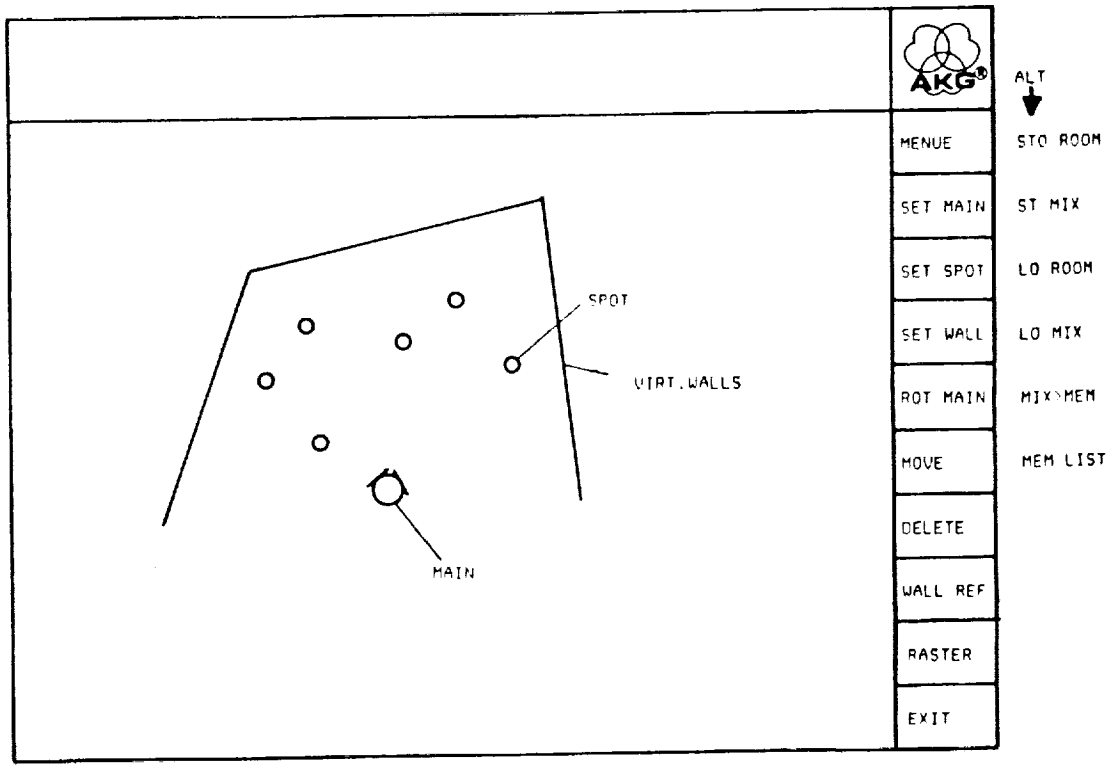


Figure 2. Illustration of the graphical interface of AKG's Creative Audio Processor for specifying characteristics of the binaural mix for a collection of independently-positioned musical instruments (adapted from product literature for the CAP 340 M).

Gary Kendall and his colleagues at Northwestern University have also been working on a real time system aimed at spatial room modeling for recording and entertainment (Kendall and Martens, 1984). Recently, Gehring Research has offered a software application for anechoic simulation using a Motorola 56001-based DSP card which uses two sets of HRTFs with the filters truncated to conform to the limitations of the DSP chip. One set is from a KEMAR manikin measured by Kendall's group and the other is from an individual subject measured by Wightman at the University of Wisconsin, Madison.

THE NASA AMES 3-D AUDITORY DISPLAY PROJECT

Since 1986, our group at NASA Ames has been working on a real time system for use in both basic research in human sound localization and applied studies of acoustic information display in advanced human-computer interfaces. The research began as part of the Ames VIrtual Environment Workstation (VIEW) project (Fisher et al., 1988). To achieve our objective, we have taken a four-part approach: (1) develop a technique for synthesizing localized, acoustic stimuli based on psychoacoustic principles, (2) in parallel, develop the signal-processing technology required to implement the synthesis technique in real time, (3) perceptually validate the synthesis technique with basic psychophysical studies, and (4) use the real time device as a research tool for evaluating and refining the approach to synthesis in both basic and applied contexts. The research has been a collaborative effort between myself as project director, Scott Foster of Crystal River Engineering (Groveland, Calif.), Fred Wightman and Doris Kistler of the University of Wisconsin, Madison, and since 1988, Durand Begault and Philip Stone at NASA Ames.

As noted above, one technique for capturing both pinnae and interaural difference cues involves binaural recording with microphones placed in the ears of a manikin (Plenge, 1974; Doll et al., 1986) or the ear canals of a human (Butler and Belendiuk, 1977). When stimuli recorded this way are presented over headphones, there is an immediate and veridical perception of 3-D auditory space (Plenge, 1974; Butler and Belendiuk, 1977; Blauert, 1983; Doll et al., 1986). Our procedure is closely related to binaural recording. Rather than record stimuli directly, we measure the acoustical transfer functions, from free-field to eardrum, at many source positions, and use these transfer functions as the basis of filters with which we synthesize stimuli. These Head-Related Transfer Functions (HRTFs), in the form of Finite Impulse Responses (FIRs), are measured using techniques adapted from Mehrgardt and Mellert (1977) (see fig. 3). Small probe microphones are placed near each eardrum of a human listener who is seated in an anechoic chamber (Wightman and Kistler, 1989a). Wide-band test stimuli are presented from 144 equidistant locations in the anechoic chamber. A new pair of impulse responses is then measured for each location in the spherical array at intervals of 15° in azimuth and 18° in elevation. HRTFs are estimated by deconvolving the loudspeakers, test stimulus, and microphone responses from the recordings made with the probe microphones (Wightman and Kistler, 1989a). The advantage of this technique is that it preserves the complex pattern of interaural differences over the entire spectrum of the stimulus, thus capturing the effects of filtering by the pinnae, head, shoulders, and torso.

For example, the insets in figure 3 show a pair of FIR filters measured for one subject for a speaker location directly to the left and at ear level, that is, at –90° in azimuth and 0° in elevation. As

you would expect, the waveform from this source arrived first and was larger in the left ear than the response measured in the right ear. The frequency-dependent effects can be analyzed by applying the Fourier Transform to these temporal waveforms.

Figure 4 shows how interaural amplitude and phase (or equivalently time) varies as a function of frequency for four different locations in azimuth at 0° in elevation. For example, the top-left panels show that for 0° in azimuth or directly in front of the listener, there is very little difference in the amplitude or phase responses between the two ears. On the other hand, in the top-right panels for 90° or directly to the listener's right, one can see that, across the frequency spectrum, the amplitude and phase responses for the right ear are larger and lead in time (phase) with respect to the left ear.

In order to synthesize localized sounds, a map of "location filters" is constructed from all 144 pairs of FIR filters by first transforming them to the frequency domain, dividing out the spectral effects of the headphones using Fourier techniques, and then transforming back to the time domain.

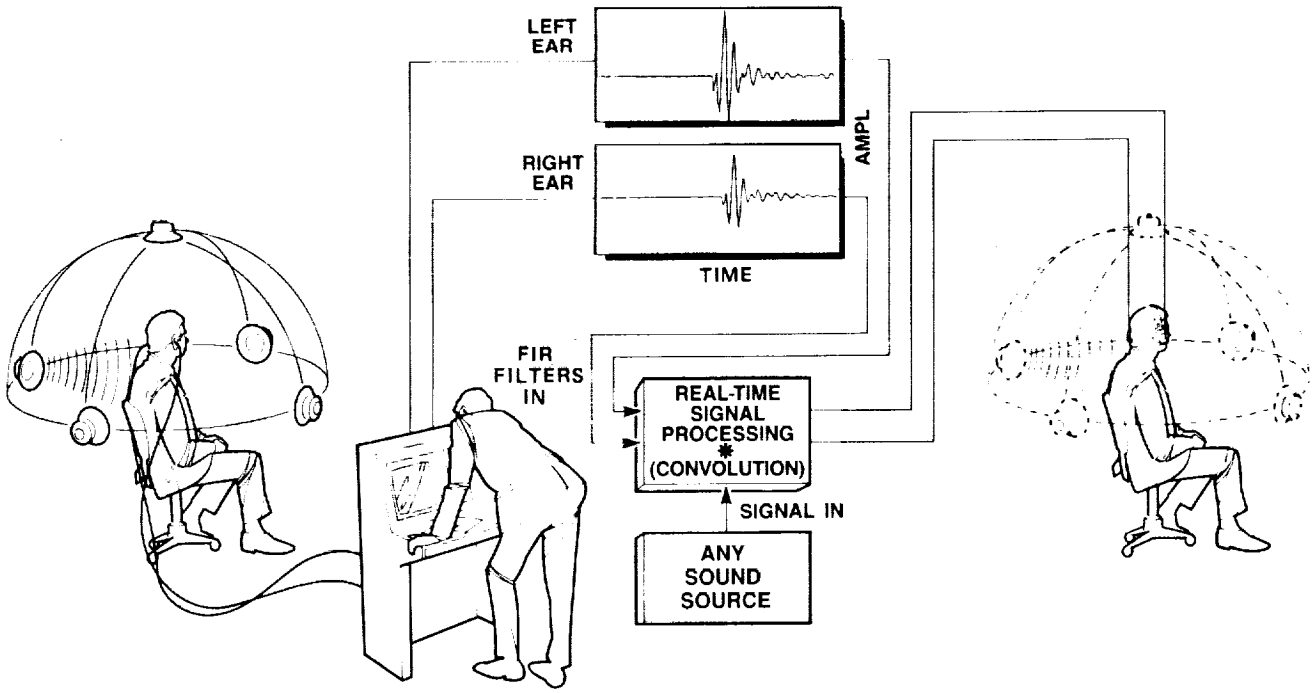


Figure 3. Illustration of the technique for synthesizing virtual acoustic sources with measurements of the head-related transfer function. An example of a pair of finite impulse responses measured for a source location at –90° to the left and 0° elevation (at ear level) is shown in the insets for the left and right ears.

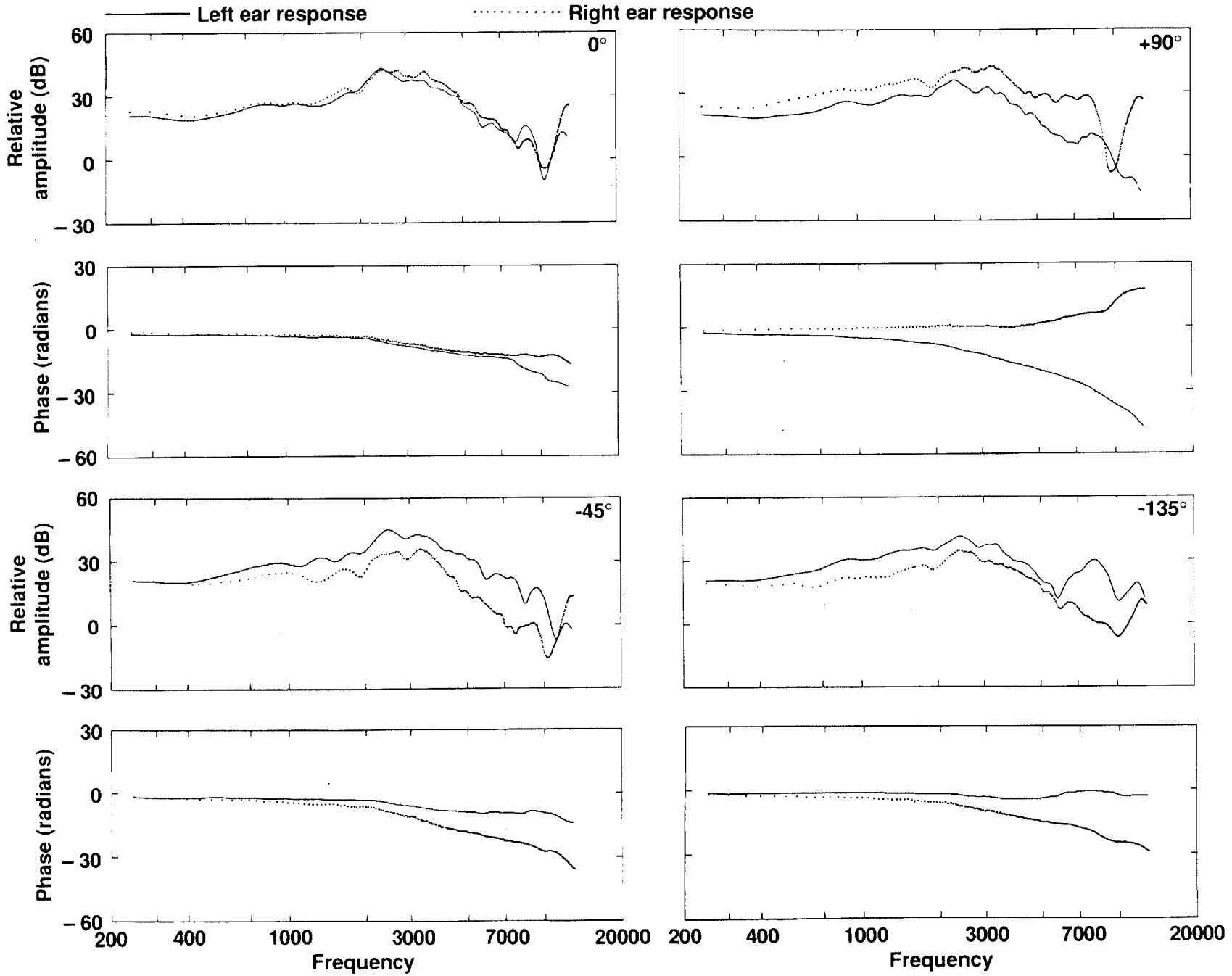


Figure 4. Examples of HRTFs in a frequency domain representation based on the Fourier transform of pairs of FIRs. Magnitude and phase responses are plotted as a function of frequency for the two ears of a single subject. Four different azimuths (0°, –45°, –135°, and +90°) at 0° elevation are shown.

The Real Time System: The Convolvotron

In the real time system, designed by Scott Foster of Crystal River Engineering, the map of corrected FIR filters is downloaded from an 80286- or 80386-based host computer to the dual-port memory of a real time digital signal-processor known as the Convolvotron (fig. 5). This set of two printed-circuit boards converts one or more monaural analog inputs to digital signals at a rate of 50 kHz (16-bit resolution). Each data stream is then convolved with filter coefficients determined by the coordinates of the desired target locations and the position of the listener's head, thus "placing" each input signal in the perceptual 3-space of the listener. The resulting data streams are mixed, converted to left and right analog signals, and presented over headphones. The current configuration allows up to four independent and simultaneous sources with an aggregate computational speed of more than 300 million multiply-accumulates per second. This processing speed is sufficient for simulating relatively small reverberant environments, and the hardware can be scaled upward to accommodate the longer filter lengths required for larger enclosures.

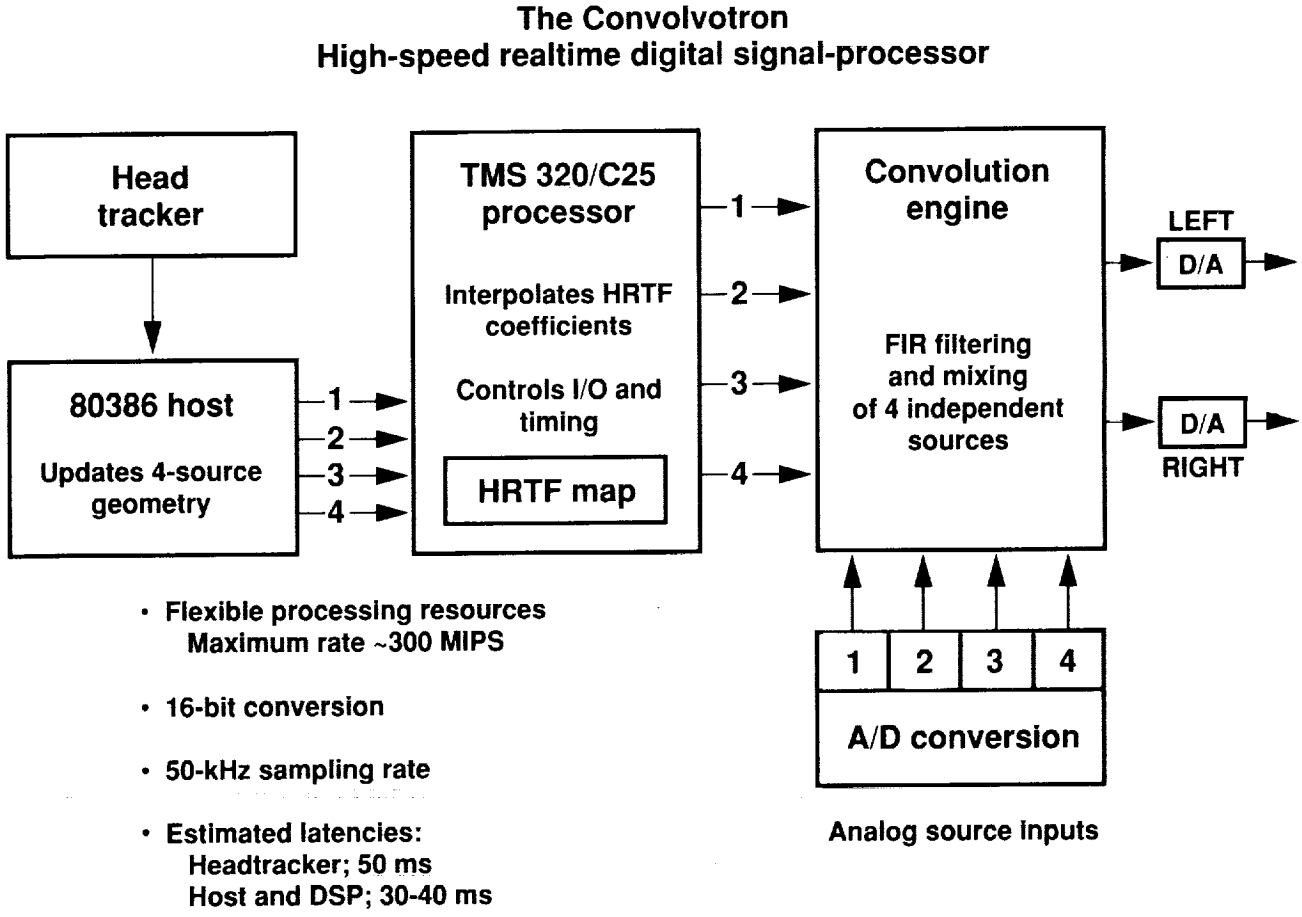


Figure 5. Block diagram of the Convolvotron system designed by Scott Foster for synthesizing three-dimensional virtual acoustic displays in real time.

Motion trajectories and static locations at greater resolution than the empirical measurements are simulated by selecting the four measured positions nearest to the target location and interpolating with linear weighting functions. The interpolation algorithm effectively computes a new coefficient at the sampling interval (every 20 μsec) so that changes in position are free from artifacts such as clicks or switching noises. When integrated with the magnetic head-tracking system (Polhemus 3-Space Isotrack), the listener's head position can be monitored in real time so that the four simultaneous sources are stabilized in fixed locations or in motion trajectories relative to the user. Such head-coupling should help to enhance the simulation since previous studies suggest that head movements are important for localization (e.g., Wallach, 1940; Thurlow et al., 1967; Thurlow and Runge, 1967). This degree of interactivity, especially coupled with simulations of simple reverberant environments, is apparently unique to the Convolvotron system.

Pilot studies at Wisconsin suggest that the interpolation approach is perceptually-viable; simple two-way linear interpolations between locations as far apart as 60° in azimuth are perceptually indistinguishable from stimuli synthesized from measured coefficients while, for elevation, localization performance begins to degrade at separations of 36°. These data suggest that the HRTF map of a real time display could tolerate interpolation separations of as much as 60° in azimuth (currently a maximum of 45° in the Convolvotron) but that the resolution of the map in elevation should probably be smaller than 36° (18° in the Convolvotron). More comprehensive evaluations of the perceptual consequences of interpolation are underway at NASA Ames.

As with any system required to compute data "on the fly," the term real time is a relative one. The Convolvotron, including the host computer, has a computational delay of about 30-40 msec, depending upon such factors as the number of simultaneous sources, the duration of the HRTFs used as filters, and the complexity of the source geometry. An additional latency of at least 50 msec is introduced by the head-tracker. This accumulation of computational delays has important implications for how well the system can simulate realistic moving sources or realistic head-motion. At the maximum delay the system can only update to a new location every 90 msec. The directional update interval, in turn, corresponds to an angular resolution of about 32° or greater when the relative source-listener speed is 360 deg/msec, 16° or greater at 180 deg/sec, and so on. Such delays may or may not result in a perceptible lag, depending upon how sensitive humans are to changes in angular displacement (the minimum audible movement angle) for a given source velocity. Recent work on the perception of auditory motion by Perrott and others using real sound sources (moving loudspeakers) suggests that these computational latencies are acceptable for moderate velocities. For example, for source speeds ranging from 8 to 360 deg/sec, minimum audible movement angles ranged from about 4 to 21°, respectively, for a 500-Hz tone-burst (Perrott, 1982; Perrott and Tucker, 1988). Thus, slower relative velocities are well within capabilities of the Convolvotron, while speeds approaching 360 deg/sec should begin to result in perceptible delays, especially when multiple sources or larger filters (e.g., simulation of simple reverberant rooms) are being generated.

Currently, the Convolvotron is being used in a variety of other government, university, and industry research labs besides ours, including the NASA Ames Crew Station Research and Development Facility, the Psychoacoustics Lab at the Research Laboratory of Electronics at MIT directed by Durlach, and Bellcore (Ludwig et al., 1990). The system also forms part of VPL Research's "Audiosphere" component of their virtual reality system.

PSYCHOPHYSICAL VALIDATION OF THE SYNTHESIS TECHNIQUE

The working assumption of our synthesis technique is that if, using headphones, we could produce ear canal waveforms identical to those produced by a free-field source, we would duplicate the free-field experience. Presumably, synthesis using individualized HRTFs would be the most likely to replicate the free-field experience for a given listener. The only conclusive test of this assumption must come from psychophysical studies in which free-field and synthesized, free-field listening are directly compared.

Validation for Static Sources Using Individualized HRTFs

A recent study by Wightman and Kistler (1989b) confirmed the perceptual adequacy of the basic approach for static sources. The stimuli were spectrally-scrambled noisebursts transduced either by loudspeakers in an anechoic chamber or by headphones. In both free-field and headphone conditions, the subjects indicated the apparent spatial position of a sound source by calling out numerical estimates of azimuth and elevation (in degrees) using a modified spherical coordinate system. For example, a sound heard directly in front would produce a response of "0, 0," a sound heard directly to the left and somewhat elevated might produce "–90 azimuth, +15 elevation," while one far to the rear on the right and below might produce "+170 azimuth, –30 elevation." Subjects were blindfolded and no feedback was given. Detailed explanations of the procedure and results can be found in the original paper.

The data analysis of localization experiments is complicated by the fact that the stimuli and responses are represented by points in three-dimensional space; in particular, as points on the surface of a unit-sphere since distance remained constant in this experiment. For these spherically-organized data, the usual statistics of means and variances are potentially misleading. For example, an azimuth error of 15° on the horizontal plane is much larger in terms of absolute distance than a 15° error at an elevation of 54°. Thus, it is more appropriate to apply the techniques of spherical statistics to characterize these psychophysical data (Fisher et al., 1987). The spherical statistic used here, the judgement centroid, is a unit-length vector with the same direction as the resultant, the vector sum of all the unit-length judgement vectors. The direction of the centroid, described by an azimuth and an elevation, can be thought of as the "average direction" of a set of judgements from the origin, the subject's position. Two indicators of variability, K^{-1} and the average angle of error, were also computed. These results will not be discussed here; the reader is referred to the original paper.

Another type of error, observed in nearly all localization studies, is the presence of front-back "confusions." These are responses which indicate that a source in the front hemisphere, usually near the median plane, is perceived to be in the rear hemisphere. Occasionally, the reverse situation is also found. It is difficult to weight these types of errors accurately. Since the confusion rate is often low (e.g., Oldfield and Parker, 1984a), reversals have generally been resolved when computing descriptive statistics; that is, the responses are coded as if the subjects had indicated the correct hemisphere, as in the analyses of table 1 and figure 6. Otherwise, estimates of error would be greatly inflated. On the other hand, if we assume that subjects' responses correctly reflect their

Table 1. Summary statistics comparing resolved localization judgements of free-field (boldface type) and virtual sources (in parentheses) for 8 subjects. (Adapted from Wightman and Kistler, 1989b)

ID	Goodness of fit	Azimuth correlation	Elevation correlation	Percent front-back reversals
SDE	**0.93** (0.89)	**0.98** (0.97)	**0.68** (0.43)	**12** (20)
SDH	**0.95** (0.95)	**0.96** (0.95)	**0.92** (0.83)	**5** (13)
SDL	**0.97** (0.95)	**0.98** (0.98)	**0.89** (0.85)	**7** (14)
SDM	**0.98** (0.98)	**0.98** (0.98)	**0.94** (0.93)	**5** (9)
SDO	**0.96** (0.96)	**0.99** (0.99)	**0.94** (0.92)	**4** (11)
SDP	**0.99** (0.98)	**0.99** (0.99)	**0.96** (0.88)	**3** (6)
SED	**0.96** (0.95)	**0.97** (0.99)	**0.93** (0.82)	**4** (6)
SER	**0.96** (0.97)	**0.99** (0.99)	**0.96** (0.94)	**5** (8)
Mean				**5.6** (11)

perceptions, resolving such confusions could be misleading. Thus, the rate of confusions is usually reported as a separate statistic.

Here, table 1 provides a general overview of the results of Wightman and Kistler (1989b). Summary statistics comparing the eight subjects' resolved judgements of location for real (free-field) and synthesized stimuli are shown; the numbers in bold-faced type are for the free-field data and the numbers in parentheses are for the synthesized conditions. Note that overall goodness of fit between the actual and estimated source co-ordinates is quite comparable, 0.89 or better for the synthesized stimuli and 0.93 or better for free-field sources. The two correlation measures indicate that while source azimuth appears to be synthesized nearly perfectly, synthesis of source elevation is more problematic, particularly for SDE who also has difficulty judging elevation in the free field. Examples of the range of patterns of localization behavior for resolved judgements can be seen in figure 6. Actual source azimuth (and, in the insets, elevation) versus the judged azimuth are plotted for subjects SDO and SDE of Wightman and Kistler (1989b). The panel on the left plots free-field judgements and the panel on the right shows judgements for the stimuli synthesized from the subjects' own transfer functions. On each graph, the positive diagonal, or a straight line with a slope of 1.0, corresponds to perfect performance.

The confusion rates (table 1) were relatively low, with average rates of about 6 and 11% for free-field and synthesized sources, respectively. Similar to the location judgements, reversal rates for the synthesized stimuli tended to be greatest for subjects who also had higher rates in the free field. Thus, while individual differences do occur, the pattern of results across synthesized and free-field conditions is consistent for a given subject; it appears that Butler and Belendiuk's (1977) observation of "good" and "bad" localizers is supported by these data.

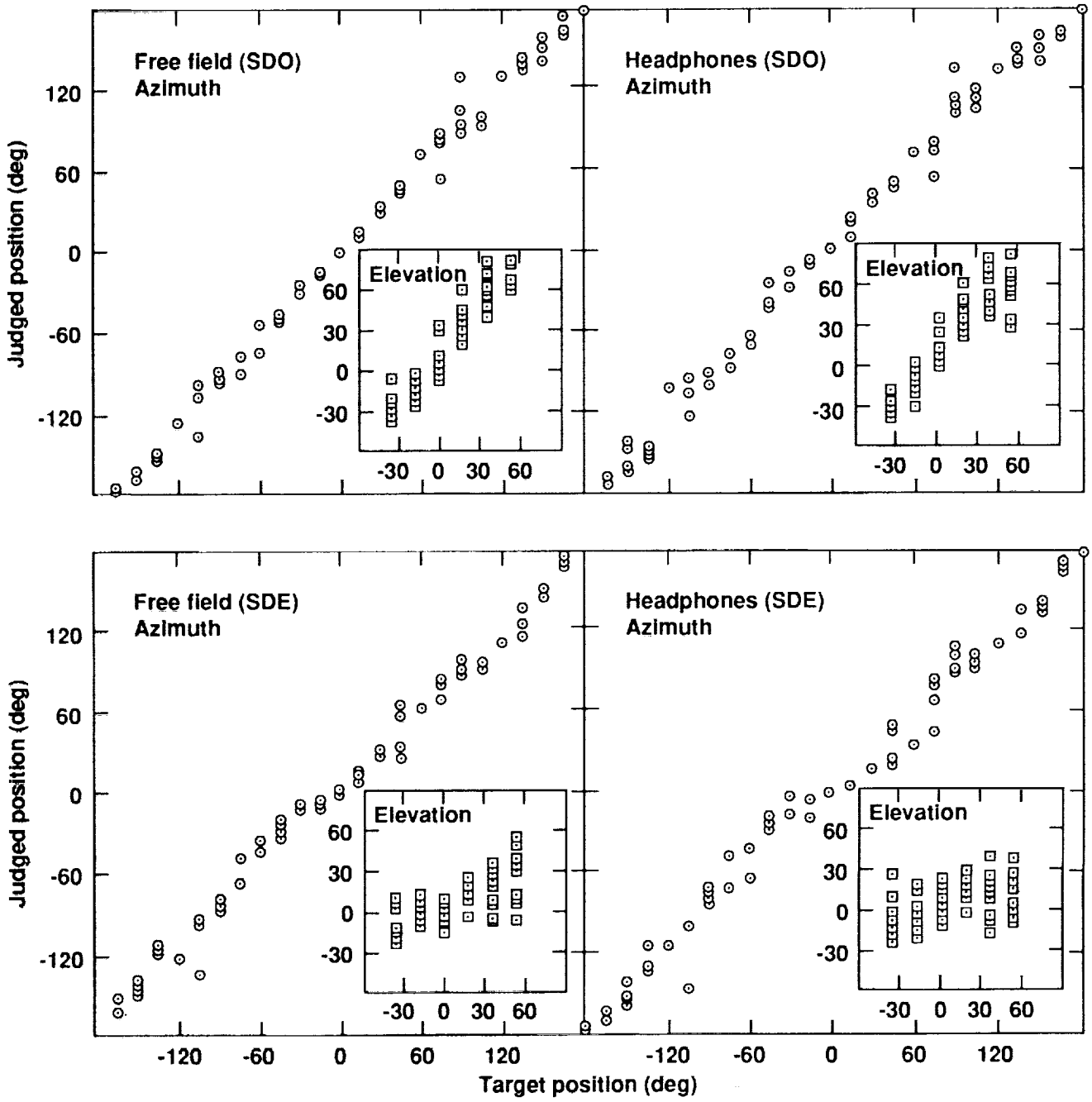


Figure 6. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subjects SDO and SDE in both free-field and headphone conditions. The plot on the left plots free-field judgements and the plot on the right shows judgements for the stimuli synthesized from the subjects' own transfer functions. Each data point represents the centroid of at least 6 judgements. 72 source positions are plotted in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 24 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots. (After Wightman and Kistler, 1989b.)

Acoustic Determinants of Performance

Individual differences in localization behavior suggest that there may be acoustic features peculiar to each subject's HRTFs which influence performance. Thus, the use of averaged transforms, or even measurements derived from normative manikins such as the KEMAR, may or may not be an optimum approach for simulating free-field sounds.

For example, figure 7 illustrates the between-subjects variability in the left and right-ear magnitude responses for a single source location (after Wenzel et al., 1988a). Obviously, any straightforward averaging of these functions would tend to smooth the peaks and valleys, thus removing potentially significant features in the acoustic transforms.

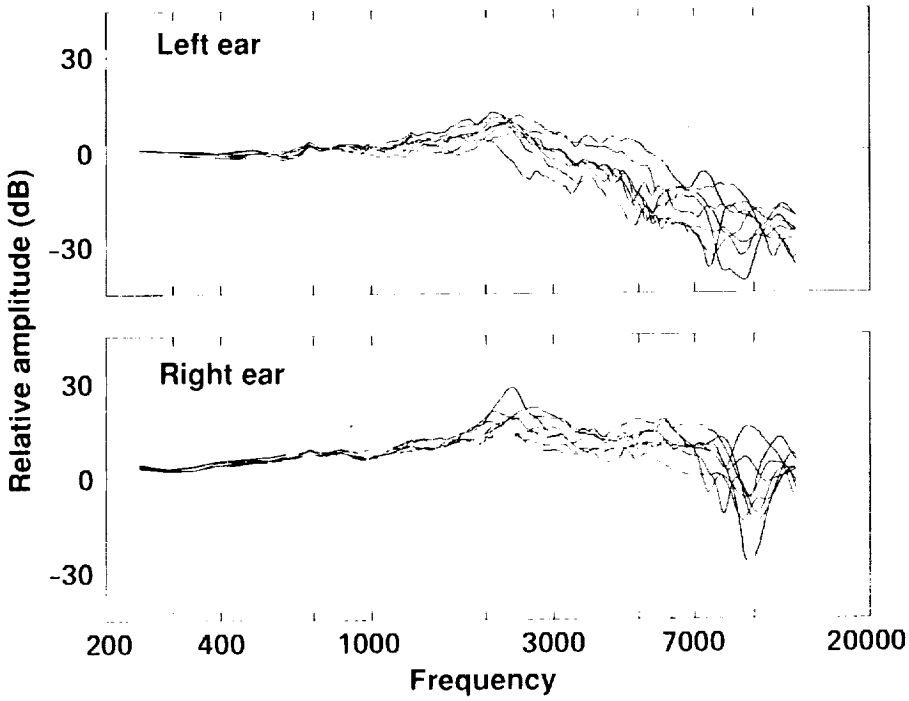


Figure 7. Magnitude responses for a single source position are shown for 8 subjects. The left and right ears are plotted separately.

On the other hand, it may be possible to identify specific features of HRTFs which result in good or bad localization. The psychophysical data indicate that elevation is particularly difficult to judge, especially for subject SDE. A preliminary analysis of elevation coding suggests that there is an acoustic basis for this poor performance.

Figure 8 plots "interaural elevation dependency" functions for four subjects' interaural amplitude data. The computational derivation of these functions can be found in the description of Wightman and Kistler's (1989b) figure 10. Essentially, the six functions on each graph show how interaural intensity changes for different elevations normalized to zero elevation, the flat function, when the magnitude responses are collapsed across all azimuths. In spite of the large intersubject variability illustrated in figure 7, the dependency functions for the better localizers (shown in the top three

graphs) are quite similar to each other and show clear elevation dependencies. SDE's functions, on the other hand, are different from the other subjects and show little change with elevation. Thus, it appears that SDE's poor performance in judging elevation for both real and synthesized stimuli may be due to a lack of distinctive acoustic features correlated with elevation in his HRTFs.

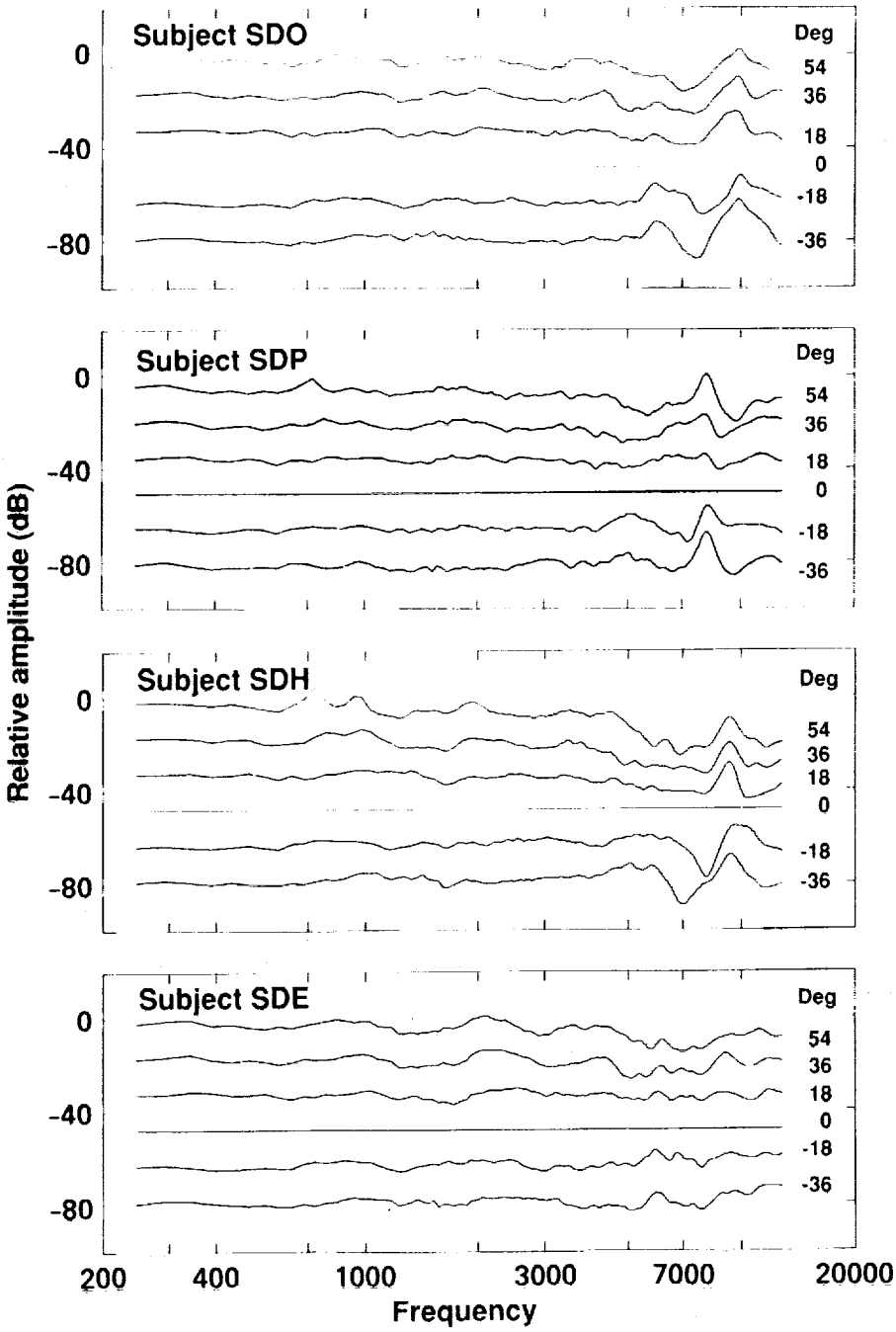


Figure 8. Interaural elevation dependency functions plotted for 4 subjects. From top to bottom, the functions within a panel represent elevations of +54, +36, +18, 0 (the reference elevation), –18, and –36°.

The analysis of individual differences in pinna cues brings up a topic which has often been conjectured about but rarely directly tested (see Butler and Belendiuk, 1977, for an early example). That is, can one manipulate localization performance simply by listening through another person's ears? Or put another way, can we adapt to and take advantage of a set of good HRTFs even if we are a bad localizer? The following data from Wenzel et al. (1988b) illustrate the kind of "cross-ear listening" paradigm that is possible using our synthesis technique. Again, the subjects provided absolute judgements of location as in the experiment by Wightman and Kistler (1989b).

Figure 9 shows what happens to resolved azimuth and elevation judgements when a good localizer listens to stimuli synthesized from another good localizer's pinna transforms. Azimuth is plotted in the top panels and elevation is on the bottom. The left and far-right graphs plot centroids for SDP's and SDO's azimuth judgements vs. the target locations when the stimuli were synthesized from their own HRTFs. Front-back confusions have been resolved as described above. As can be seen, both SDP and SDO localize the synthesized stimuli based on their own HRTFs quite well. The center graphs show what happens when SDP listens "through" SDO's pinnae. Localization of azimuth degrades somewhat, but not a great deal. Elevation performance degrades further, suggesting that elevation cues are not as robust as azimuth cues across the range of individuals, but an overall correspondence between real and perceived locations remains intact.

Figure 10 compares performance when a good localizer, SDO, listens to stimuli synthesized from the HRTFs of bad localizer SDE. Again for azimuth there is little degradation. However, for elevation, it seems that SDE's pinnae provide poor elevation cues for SDO as well, supporting the notion that acoustic features of the transforms determine localization.

If acoustic features do determine localization, one might conclude the reciprocal case is true; that SDE could actually improve his performance if he could listen "through" SDO's ears. Figure 11 plots these data. Again, SDE, whose azimuth judgements are accurate for stimuli synthesized from his own HRTFs, performs nearly as well when listening to SDO's azimuth cues. However, it appears that cross-ear listening is not a symmetrical effect for elevation. Even after about 50 hr of testing, compared to only 2 hr for the good localizers, SDE still could not take advantage of the presumably better cues provided by SDO's pinnae. These data are hardly conclusive since they are based on a sample size of one; only SDE of the eight subjects in Wightman and Kistler (1989b) showed such poor elevation performance to begin with. But they are suggestive. It may be that there is a critical period for localization which, once past, can never be regained. Perhaps more likely is that, analogous to the experiments with prisms in visual adaptation (see Welch, 1978), SDE would need prolonged and consistent exposure to SDO's pinnae in order to learn to discriminate the subtle acoustic cues he does not normally experience. Apparently, a few hours of testing a day, especially in the absence of either verbal feedback or correlated information from the other senses, are not enough to allow adaptation to occur.

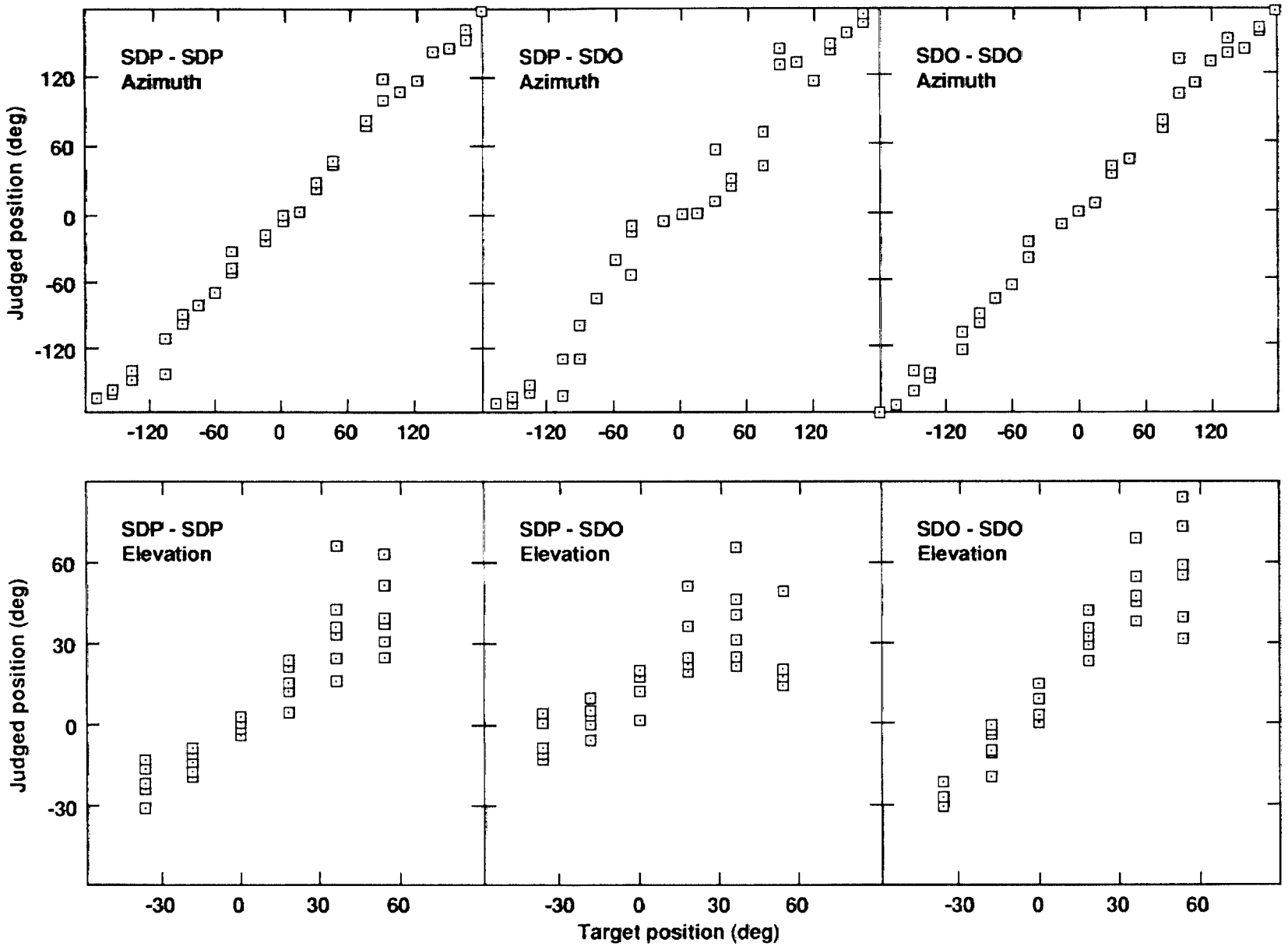


Figure 9. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for three different headphone conditions. The plots on the far left and right show SDP's and SDO's judgements for stimuli synthesized from their own transfer functions. The center plot shows SDP's judgements for stimuli synthesized from SDO's HRTFs. Each data point represents the centroid of at least 6 judgements. 36 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.

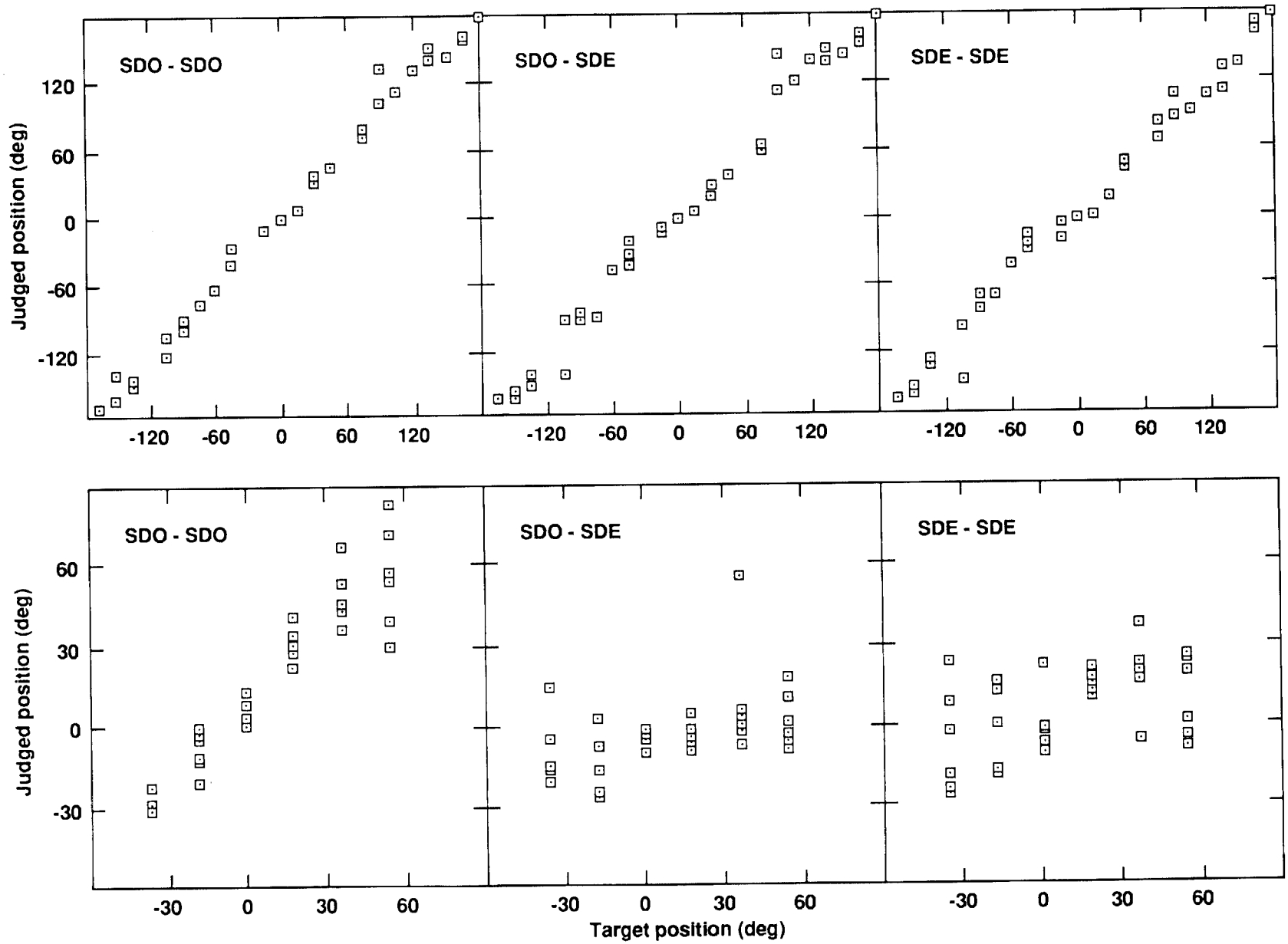


Figure 10. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for three different headphone conditions. The plots on the far left and right show SDO's and SDE's judgements for stimuli synthesized from their own transfer functions. The center plot shows SDO's judgements for stimuli synthesized from SDE's HRTFs. Each data point represents the centroid of at least 6 judgements. 36 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.

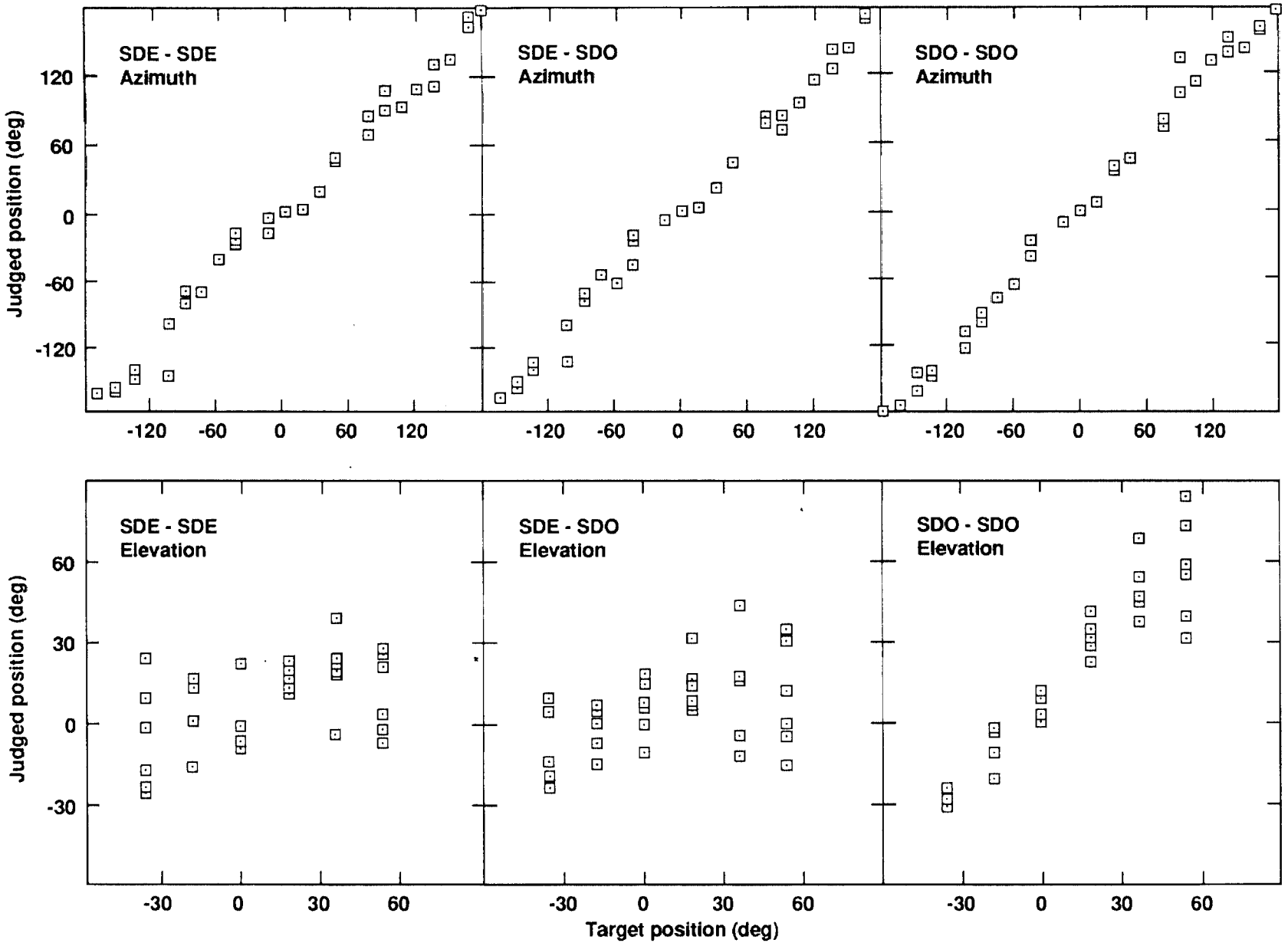


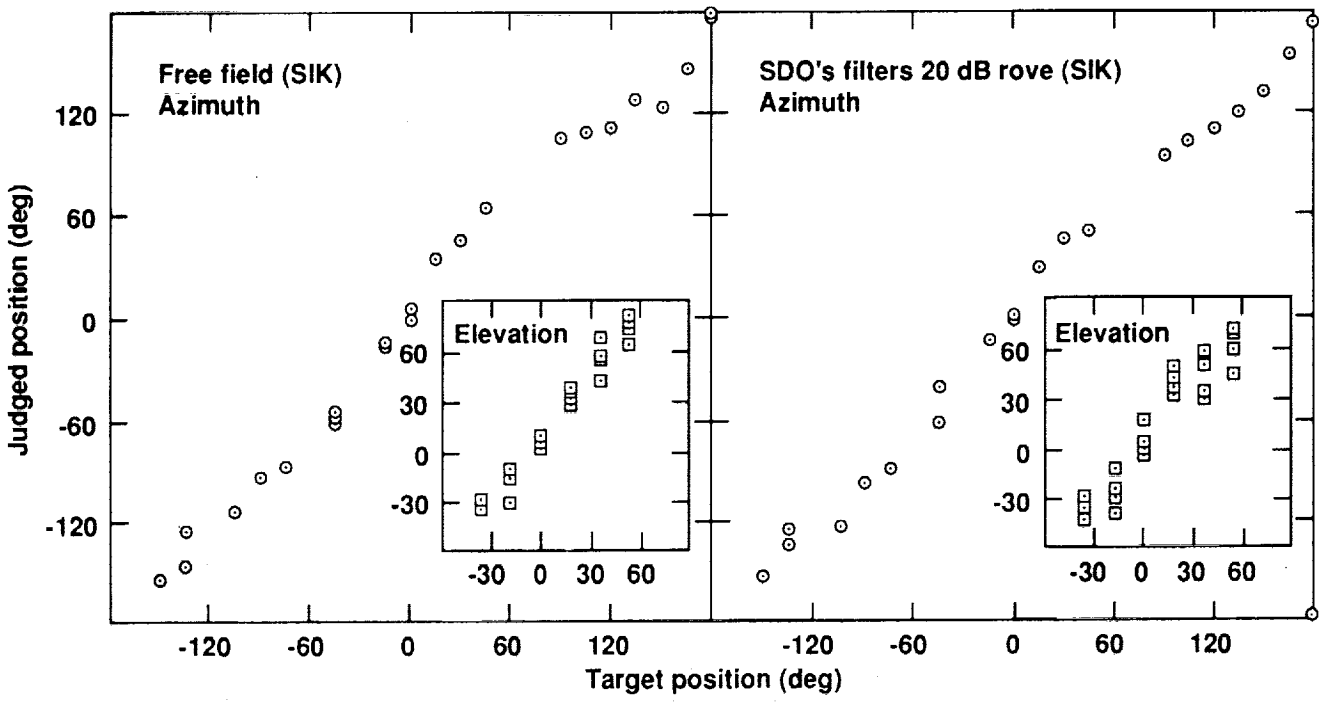
Figure 11. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for three different headphone conditions. The plots on the far left and right show SDE's and SDO's judgements for stimuli synthesized from their own transfer functions. The center plot shows SDE's judgements for stimuli synthesized from SDO's HRTFs. Each data point represents the centroid of at least 6 judgements. 36 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.

Inexperienced Listeners and Nonindividualized HRTFs

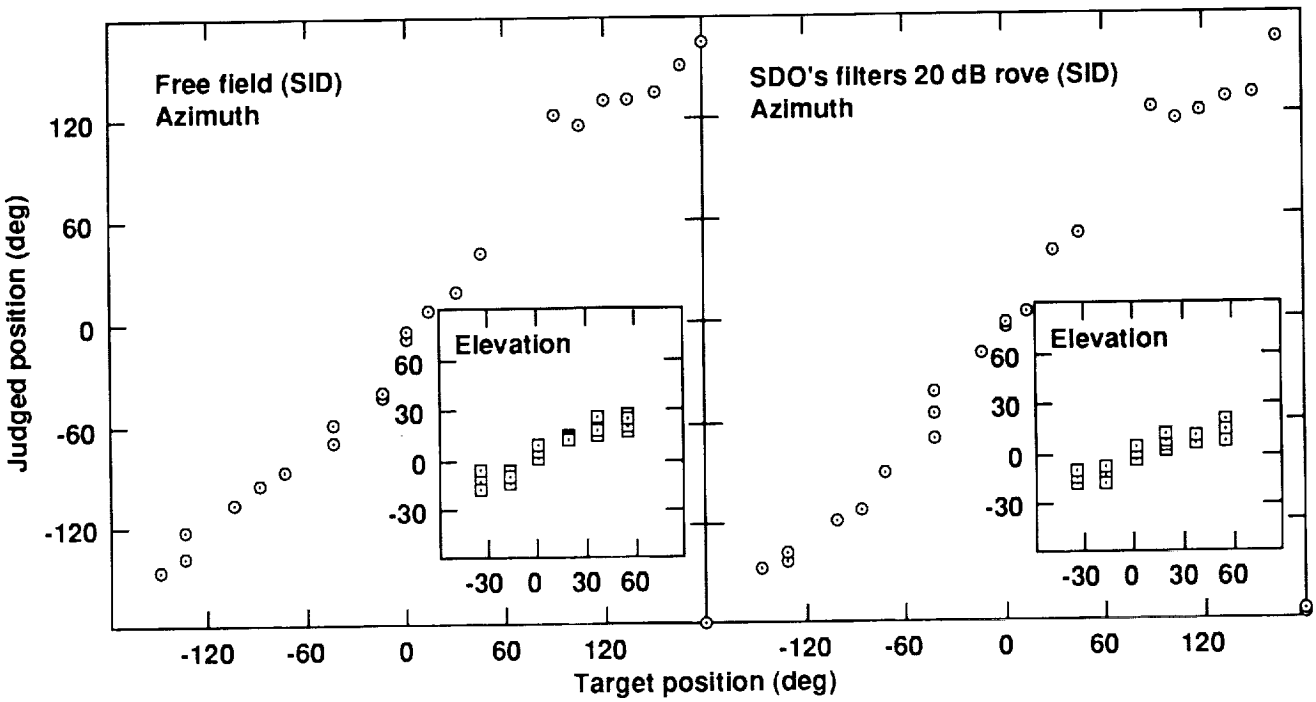
In practice, measurement of each potential listener's HRTFs may not be feasible. It may also be the case that the user of a 3D auditory display will not have the opportunity for extensive training. Thus, a critical research issue for virtual acoustic displays is the degree to which the general population of listeners can readily obtain adequate localization cues from stimuli based on nonindividualized transforms. The individual difference data of figures 9 through 11 suggest that, even in the worst case, using nonindividualized transforms does not degrade localization accuracy much more than the listener's inherent ability. In general, then, even inexperienced listeners may be able to use a particular set of HRTFs as long as they provide adequate cues for localization. A reasonable approach is to use the HRTFs from a subject whose measurements have been "behaviorally-calibrated" and are thus correlated with known perceptual ability in both free-field and headphone conditions. Recently, Wenzel et al. (1991) completed a more extensive study using a variant on the cross-ear listening paradigm; 16 inexperienced listeners judged the apparent spatial location of sources presented over loudspeakers in the free field or over headphones. The headphone stimuli were generated digitally using HRTFs measured in the ear canals of a representative subject, SDO, a "good localizer" from the experiment by Wightman and Kistler (1988b).

Figure 12 illustrates the behavior of 12 of the 16 subjects. When front-back confusions are resolved, localization performance is quite good, with judgements for the nonindividualized stimuli nearly identical to those in the free-field. Like SDE in Wenzel et al. (1988b), 2 of the subjects show poor elevation performance in both free-field and headphone conditions, a response pattern which is at least consistent across the free-field and virtual source conditions (fig. 13). The third pattern is illustrated in figure 14; here, 2 subjects show inconsistent behavior with poor elevation accuracy in only the synthesized conditions. The latter phenomenon, if it turns out to be common, would be a problem for virtual displays.

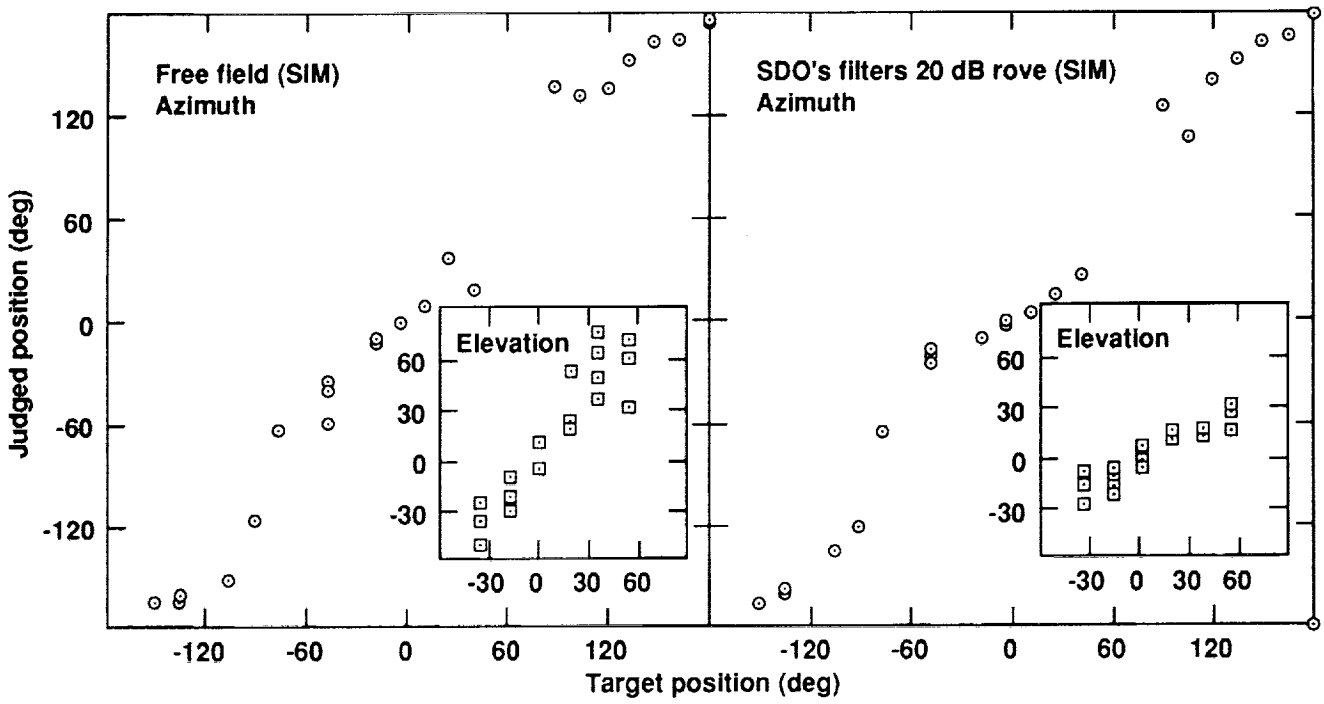
In general, these data suggest that most listeners can obtain useful directional information from an auditory display without requiring the use of individually-tailored HRTFs, particularly for azimuth. However, a caveat is important here. Again, the results plotted in figures 6 and 9 through 14 are based on analyses in which errors due to front/back confusions are resolved. For free-field versus simulated free-field stimuli, experienced listeners in the Wightman and Kistler study exhibit front/back confusion rates of about 6 vs. 11% while the inexperienced listeners show average rates of about 19 vs. 31%. Note, though, that the existence of free-field confusions indicates that these reversals are not strictly the result of the simulation. It is possible, as Asano et al. (1990) have claimed, that these errors diminish as subjects adapt to the unusual listening conditions provided by static anechoic sources, whether real or simulated. The difference in free-field confusion rates between the inexperienced listeners of this experiment and the more experienced subjects of Wightman and Kistler tend to support this view. Thus, it may be that some form of adaptation or training with feedback will be required to take full advantage of a virtual acoustic display.



Figures 12. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subject SIK in both free-field and headphone conditions. The plot on the left plots free-field judgements and the plot on the right shows judgements for the stimuli synthesized from nonindividualized transfer functions. Each data point represents the centroid of 9 judgements. 24 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.



Figures 13. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subject SID in both free-field and headphone conditions. The plot on the left plots free-field judgements and the plot on the right shows judgements for the stimuli synthesized from nonindividualized transfer functions. Each data point represents the centroid of 9 judgements. 24 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.

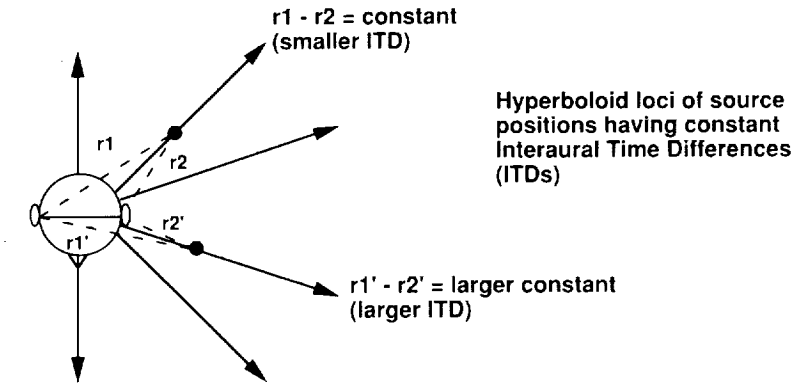


Figures 14. Scatterplots of actual source azimuth (and, in the insets, elevation) versus judged source azimuth for subject SIM in both free-field and headphone conditions. The plot on the left plots free-field judgements and the plot on the right shows judgements for the stimuli synthesized from nonindividualized transfer functions. Each data point represents the centroid of 9 judgements. 24 source positions are given in each plot. Data from 6 different source elevations are combined in the azimuth plots and data from 18 different source azimuths are combined in the elevation insets. Note that the scale is the same in the azimuth and elevation plots.

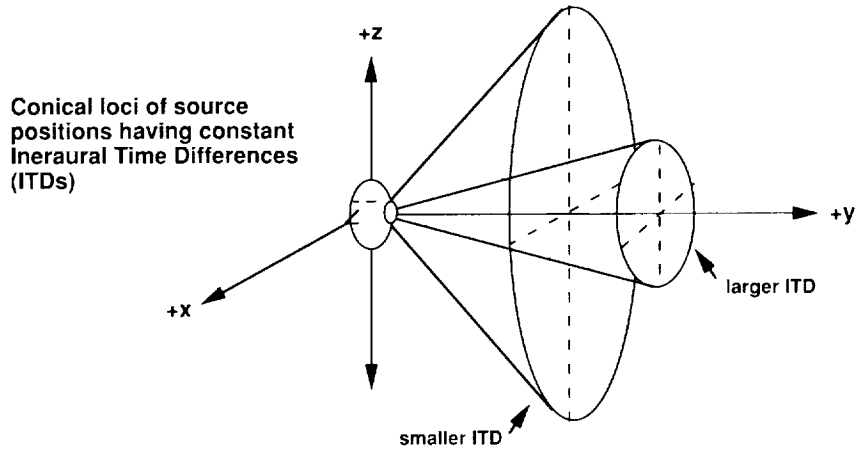
IMPROVING VIRTUAL ACOUSTIC DISPLAYS: PROBLEM AREAS AND RESEARCH ISSUES

Although the reason for errors such as front-back confusions is not completely understood, they are probably due in large part to the static nature of the stimulus and the ambiguity resulting from the so-called cone of confusion (Mills, 1972). Assuming a stationary, spherical model of the head, a given interaural time difference correlates ambiguously with the direction of a sound source, with a conical shell describing the locus of all possible sources (fig. 15). However, cone-of-confusion effects alone cannot explain a front-to-back response bias, and it may be that visual dominance plays a substantial role in auditory localization (see Warren et al., 1981). That is, given an ambiguous acoustic stimulus in the absence of an obvious visual correlate, it may be that the perceptual system resolves the ambiguity with a heuristic that assumes the source is behind the listener where it can't be seen.

Several stimulus characteristics may help to minimize these errors. For example, the addition of visual cues, dynamic cues correlated with head motion, and well-controlled environmental cues



(a) Two-dimensional, overhead view



(b) Three-dimensional view

Figure 15. Illustration of the cone-of-confusion effect for different interaural delays. Assuming a spherical head and symmetrically-located ear canals, all sound sources lying along a hyperbolic surface would produce the same interaural delay in two dimensions (e.g., the horizontal plane) and a conical surface in three dimensions.

derived from models of room acoustics may improve the ability to resolve these ambiguities. By taking advantage of the head-tracker in the real time system, we can close the loop between the auditory, visual, vestibular, and kinesthetic systems and study the effects of dynamic interaction with relatively complex, but known, acoustic environments.

A related problem in synthesizing veridical acoustic images is the fact that such stimuli sometimes fail to externalize, particularly when the signals are unfamiliar (e.g., the spectrally-scrambled noisebursts used here) and simulated from anechoic measurements of HRTFs. Thus cues which provide a sense of distance and environmental context, such as the ratio of direct to reflected energy and other characteristics specific to particular enclosed spaces, may also enhance the externalization of images (Coleman, 1963; Gardner, 1968; Laws, 1972; 1973; Plenge, 1974; Borish, 1984; Begault,

1987; 1990). Further, just as we come to learn the characteristics of a particular room or concert hall, the localization of virtual sounds may improve if the listener is allowed to become familiar with sources as they interact in a particular artificial acoustic world. For example, perhaps simulation of an asymmetric room would tend to aid the listener in distinguishing front from rear locations (Begault and Wenzel, in progress). However, the specific parameters used in such a model must be investigated carefully if localization accuracy is to remain intact. For example, Blauert (1983) reports that the spatial image of a sound source grows larger and increasingly diffuse with increasing distance in a reverberant environment, a phenomenon which may tend to interfere with the ability to judge the direction of the source. Further, the success of any reasonably-complex spatial display will depend upon our understanding of localization masking, or the stimulus parameters which affect the identification, segregation (e.g., Bregman, 1990), and discrimination (e.g., Perrott, 1984a,b) of multiple sources. Surprisingly, little or no research has been done on the localization of more than two simultaneous sources.

Another critical area for research is the further specification of the role of individual differences and perhaps the development of efficient techniques for training or adaptation to nonindividualized transforms. The fact that individual differences in performance are apparently correlated with acoustical idiosyncrasies in the HRTFs suggests that the systematic analysis and manipulation of HRTF characteristics may provide a means for counteracting individual difference effects. Given appropriate adaptation techniques, it may eventually be possible to construct a set of "universal transforms" using parametric techniques like Genuit's structural model (1986), data reduction techniques like specialized averaging models and principal components analysis (Asano et al., 1990; Kistler and Wightman, 1990), or perhaps even enhancing the features of empirically-derived transfer functions (Durlach and Pang, 1986).

Other research will be related to further refinements in the techniques for the measurement, manipulation, and perceptual validation of HRTFs, including practical signal-processing issues such as determining optimal techniques for interpolation between measured or modeled transforms to ensure veridical motion.

The simulation techniques investigated here provide both a means of implementing a virtual acoustic display and the ability to study features of human sound localization that were previously inaccessible due to a lack of control over the stimuli. The availability of real time control systems (e.g., Wenzel et al., 1988a) further expand the scope of the research, allowing the study of dynamic, intersensory aspects of localization which may do much toward alleviating the problems encountered in producing the reliable and veridical perception which is critical for many applied contexts.

REFERENCES

Asano, F.; Suzuki, Y.; and Sone, T.: Role of spectral cues in median plane localization. J. Acoust. Soc. Am., vol. 88, 1990, pp. 159-168.

Begault, D. R.: (1987) Control of Auditory Distance. Dissertation, University of California, San Diego, Calif.

Begault, D. R.: Challenges to the successful implementation of 3-D Sound. Proceedings of the 89th Meeting of the Audio Engineering Society, Los Angeles, Calif., Sept. 21-25, 1990.

Begault, D. R.; and Wenzel, E. M.: Techniques and applications for binaural sound manipulation in man-machine interfaces. NASA TM-102279, 1990.

Blattner, M. M.; Sumikawa, D. A.; and Greenberg, R. M.: Earcons and Icons: Their Structure and Common Design Principles. Human-Computer Interaction, vol. 4, 1989, pp. 11-44.

Blauert, J.: Sound localization in the median plane. Acustica, vol. 22, 1969, pp. 205-213.

Blauert, J.: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press: Cambridge, Mass., 1983.

Blauert, J.: Psychoakustik des binauralen Horens. The psychophysics of binaural hearing. Invited plenary paper, DAGA'84, Darmstadt, FRG, 1984.

Bly, S.: Sound and computer information presentation. Doctoral thesis (UCRL-53282) Lawrence Livermore National Laboratory and University of California, Davis, Calif., 1982.

Boerger, G.; Laws, P.; and Blauert, J.: Stereophonic headphone reproduction with variation of variation of various transfer factors by means of rotational head movements. Acustica, vol. 39, 1977, pp. 22-26.

Borish, J.: Extension of the image model to arbitrary polyhedra. J. Acoust. Soc. Amer., vol. 75, 1984, pp. 1827-1836.

Bregman, A. S.: Asking the "What for" question in auditory perception. In Perceptual Organization, M. Kubovy and J. R. Pomerantz, eds., Lawrence Erlbaum Associates (Hillsdale, NJ), 1981.

Bregman, A. S.: Auditory Scene Analysis. MIT Press (Cambridge, Mass.), 1990.

Bronkhorst, A. W.; and Plomp, R.: The effect of head-induced interaural time and level differences on speech intelligibility in noise. J. Acoust. Soc. Amer., vol. 83, 1988, pp. 1508-1516.

Brooks, F. P.: Grasping reality through illusion—Interactive graphics serving science. Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys., Washington, DC, 1988, pp. 1-11.

Butler, R. A.; and Belendiuk, K.: Spectral cues utilized in the localization of sound in the median sagittal plane. J. Acoust. Soc. Am., vol. 61, 1977, pp. 1264-1269.

Buxton, W.; Gaver, W.; and Bly, S.: The use of non-speech audio at the interface. Tutorial #10, CHI'89, ACM Press (New York), 1989.

Calhoun, G. L.; Valencia, G.; and Furness, T. A. III: Three-dimensional auditory cue simulation for crew station design/evaluation. Proc. Hum. Fac. Soc., vol. 31, 1987, pp. 1398-1402.

Carterette, E.; and Friedman, M.; eds.: Hearing, Handbook of Perception (Volume IV). Academic Press (New York), 1978.

Cherry, E. C.: Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am., vol. 22, 1953, pp. 61-62.

Coleman, P. D.: An analysis of cues to auditory depth perception in free space. Psych. Bull., vol. 60, 1963, pp. 302-315.

Colquhoun, W. P.: Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. Hum. Fac., vol. 17, 1975, pp. 425-437.

Cooper, D. H.; and Bauck, J. L.: Prospects for transaural recording. J. Aud. Eng. Soc., vol. 37, 1989, pp. 3-19.

Deatherage, B. H.: Auditory and other sensory forms of information presentation. In H. P. Van Cott and R. G. Kincade, eds., Human Engineering Guide to Equipment Design, (rev. ed.), U.S. Government Printing Office, 1972, pp. 123-160.

Deutsch, D.; ed.: The Psychology of Music. Academic Press (New York), 1982.

Doll, T. J.; Gerth, J. M.; Engelman, W. R.; and Folds, D. J.: Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014, 1986.

Durlach, N. I.; and Pang, X. D.: Interaural magnification. J. Acoust. Soc. Am., vol. 80, 1986, pp. 1849-1850.

Durlach, N. I.: Auditory localization in teleoperator and virtual environment systems: Ideas, issues, and problems. Perception (in press).

Durlach, N. I.: (1991) Auditory localization in teleoperator and virtual environment systems: Ideas, issues, and problems. Perception (in press).

Edwards, A. D. N.: Soundtrack: An auditory interface for blind users. Hum. Comp. Interact., vol. 4, 1989, pp. 45-66.

Fisher, N. I.; Lewis, T.; and Embleton, B. J. J.: Statistical Analysis of Spherical Data. Cambridge U. Press.: Cambridge, U.K., 1987.

Fisher, S. S.; Wenzel, E. M.; Coler, C.; and McGreevy, M. W.: Virtual interface environment workstations. Proc. Hum. Fac. Soc., vol. 32, 1988, pp. 91-95.

Foley, J. D.: Interfaces for advanced computing. Sci. Amer., vol. 257, 1987, pp. 126-135.

Forbes, T. W.: Auditory signals for instrument flying. J. Aeronaut. Soc., 1946, May, pp. 255-258.

Furness, T. A.: The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., vol. 1, 1986, pp. 48-52.

Garner, W. R.: Auditory signals. In A Survey Report on Human Factors in Undersea Warfare, Nat. Res. Council, Washington, DC, 1949, pp. 201-217.

Gardner, M. B.: Proximity image effect in sound localization. J. Acoust. Soc. Am., vol. 43, 1968, p. 163.

Gardner, M. B.; and Gardner, R. S.: Problem of localization in the median plane: Effect of pinnae cavity occlusion. J. Acoust. Soc. Am., vol. 53, 1973, pp. 400-408.

Gaver, W. W.: Auditory icons: Using sound in computer interfaces. Hum.-Comp. Interact., vol. 2, 1986, pp. 167-177.

Gaver, W. W.; and Smith, R. B.: Auditory icons in large-scale collaborative environments. Human Computer Interaction – Proceedings of Interact'90. Elsevier (North Holland), 1990.

Genuit, K.: A description of the human outer ear transfer function by elements of communication theory. Proc. 12th ICA (Toronto), Paper B6-8, 1986.

Gibson, J. J.; The ecological approach to visual perception. Houghton Mifflin (Boston), 1979.

Gierlich, H. W.; and Genuit, K.: Processing artificial-head recordings. J. Aud. Eng. Soc., vol. 37, 1989, pp. 34-39.

Hudde, H.; and Schroter, J.: (Improvements in the Neumann artificial head system). In Runkfunktechnische Mitteilungen (Radio Technology Reports), FRG, 1981.

Kendall, G. S.; and Martens, W. L.: Simulating the cues of spatial hearing in natural environments. Proceedings of the 1984 International Computer Music Conference, Paris, 1984.

Kendall, G. S.; and Wilde, M. D.: Production and reproduction of three-dimensional spatial sound. 87th Meeting of the Audio Engineering Society, J. Aud. Eng. Soc. (Abstracts), vol. 37, 1989, p. 1066.

Kistler, D. J.; and Wightman, F. L.: Principal components analysis of head-related transfer functions. J. Acoust. Soc. Am., vol. 88, 1990, S98.

Kubovy, M.; and Howard, F. P.: Persistence of a pitch-segregating echoic memory. J. Exp. Psych: Hum. Perc. Perf., vol. 2, 1976, pp. 531-537.

Kubovy, M.: Concurrent pitch-segregation and the theory of indispensable attributes. In Perceptual Organization, M. Kubovy and J. R. Pomerantz, eds., Lawrence Erlbaum Associates (Hillsdale, NJ), 1981.

Laws, P.: Zum Problem des Entfernungshorens und der Im-Kopf-Lokalisiertheit von Horeregnissen. (On the problem of distance hearing and the localization of auditory events inside the head). Dissertation, Technische Hochschule, Aachen, FRG, 1972.

Laws, P.: Entfernungshoeren und das Problem der Im-Kopf-Lokalisiertheit von Hoerereignissen. (Auditory distance perception and the problem of "in-head localization of sound images), Acustica, vol. 29, 1973, pp. 243-259.

Lehnert, H.; and Blauert, J.: A concept for binaural room simulation. ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, Oct. 15-18, 1989.

Loomis, J. M.; Hebert, C.; and Cicinelli, J. G.: (1990) Active localization of virtual sounds. J. Acoust. Soc. Am., vol. 88, 1990, pp. 1757-1764.

Lord Rayleigh: (Strutt, J. W.) On our perception of sound direction. Phil. Mag., vol. 13, 1907, pp. 214-232.

Ludwig, L.; Pincever, N.; and Cohen, M.: Extending the notion of a window system to audio. Computer, 1990, pp. 66-72.

McKinley, R. L.; and Ericson, M. A.: Digital synthesis of binaural auditory localization azimuth cues using headphones. J. Acoust. Soc. Am., vol. 83, 1988, S18.

Mehrgardt, S.; and Mellert, V.: Transformation characteristics of the external human ear. J. Acoust. Soc. Am., vol. 61, 1977, pp. 1567-1576.

Mills, A. W.: Auditory localization. In Foundations of Modern Auditory Theory, Vol. II, J. V. Tobias, ed., Academic Press (New York), 1972.

Minsky, M.; Ming, O.; Steele, O.; Brooks, F. P.; and Behensky, M.: Feeling and Seeing: Issues in Force Display. Computer Graphics, vol. 24, 1990, pp. 235-243.

Mowbray, G. H.; and Gebhard, J. W.: Man's senses as informational channels. In Human Factors in the Design and Use of Control Systems. H. W. Sinaiko, ed., Dover Publications (New York), 1961.

O'Leary, A.; and Rhodes, G.: Cross-modal effects on visual and auditory object perception. Perc. and Psychophys., vol. 35, 1984, pp. 565-569.

Oldfield, S. R.; and Parker, S. P. A.: Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. Perc., vol. 13, 1984a, pp. 581-600.

Oldfield, S. R.; and Parker, S. P. A.: Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. Perc., vol. 13, 1984b, pp. 601-617.

Oldfield, S. R.; and Parker, S. P. A.: Acuity of sound localisation: a topography of auditory space. III. Monaural hearing conditions. Perc., vol. 15, 1986, pp. 67-81.

Patterson, R. R.: Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority Paper No. 82017, London, 1982.

Perrott, D. R.: Studies in the perception of auditory motion. In Localization of Sound: Theory and Applications. R. W. Gatehouse, ed., Amphora Press (Groton, CT), 1982.

Perrott, D. R.: Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity. J. Acoust. Soc. Am., vol. 75, 1984a, pp. 1201-1206.

Perrott, D. R.: Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays. J. Acoust. Soc. Am., vol. 76, 1984b, pp. 1704-1712.

Perrott, D. R.; and Tucker, J.: Minimum audible movement angle as a function of signal frequency and the velocity of the source. J. Acoust. Soc. Am., vol. 83, 1988, pp. 1522-1527.

Perrott, D. R.; Sadralodabai, T.; and Saberi, K.: Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors (In press.)

Persterer, A.: A very high performance digital audio signal processing system. ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, Oct. 15-18, 1989.

Plenge, G.: On the difference between localization and lateralization. J. Acoust. Soc. Am., vol. 56, 1974, pp. 944-951.

Posselt, C.; Schroter, J.; Opitz, M.; Divenyi, P.; and Blauert, J.: Generation of binaural signals for research and home entertainment. Proc. 12th ICA (Toronto), Paper B1-6, 1986.

Shaw, E. A. G.: The external ear. In Handbook of Sensory Physiology, Vol. V/1, Auditory System, W. D. Keidel and W. D. Neff, eds., Springer-Verlag (New York), 1974.

Shaw, E. A. G.: The external ear: New knowledge. In Earmolds and Associated Problems. S. C. Dalsgaard, ed., Proc. 7th Danavox Symposium, Scand. Audiology, Suppl., vol. 5, 1975, pp. 24-50.

Smith, S.; Bergeron. R. D.; and Grinstein, G. G.: Stereophonic and surface sound generation for exploratory data analysis. Proc. CHI'90, ACM Conf. Hum. Fac. Comp. Sys., Seattle, Wash., 1990, pp. 125-132.

Sorkin, R. D.; Wightman, F. L.; Kistler, D. J.; and Elvers, G. C.: An exploratory study of the use of movement-correlated cues in an auditory head-up display. Hum. Fac., vol. 31, 1989, pp. 161-166.

Sutherland, I. E.: Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., vol. 33, 1968, pp. 757-764.

Thurlow, W. R.; Mangels, J. W.; and Runge, P. S.: Head movements during sound localization. J. Acoust. Soc. Am., vol. 42, 1967, pp. 489-493.

Thurlow, W. R.; and Runge, P. S.: Effects of induced head movements on localization of direction of sound sources. J. Acoust. Soc. Am., vol. 42, 1967, pp. 480-488.

Wallach, H.: The role of head movements and vestibular and visual cues in sound localization. J. Exp. Psychol., vol. 27, 1940, pp. 339-368.

Warren, D. H.; Welch, R. B.; and McCarthy, T. J.: The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. Perc. and Psychophys., vol. 30, 1981, pp. 557-564.

Welch, R. B.: Perceptual Modification: Adapting to altered sensory environments. Academic Press (New York), 1978.

Wenzel, E. M.; Wightman, F. L.; and Foster, S. H.: A virtual display system for conveying three-dimensional acoustic information. Proc. Hum. Fac. Soc., vol. 32, 1988a, pp. 86-90.

Wenzel, E. M.; Wightman, F. L.; Kistler, D. J.; and Foster, S. H.: Acoustic origins of individual differences in sound localization behavior. J. Acoust. Soc. Amer., vol. 84, 1988b, S79.

Wenzel, E. M.; Stone, P. K.; Fisher, S. S.; and Foster, S. H.: A system for three-dimensional acoustic "visualization" in a virtual environment workstation. Proceedings of the IEEE Visualization '90 Conference, San Francisco, Calif., Oct. 23-26, 1990, pp. 329-337.

Wenzel, E. M.; Wightman, F. L.; and Kistler, D. J.: Localization of non-individualized virtual acoustic display cues. Proceedings of the CHI'91, ACM Conference on Computer-Human Interaction, New Orleans, La., April 27-May 2, 1991.

Wightman, F. L.; and Kistler, D. J.: Headphone simulation of free-field listening I: stimulus synthesis. J. Acoust. Soc. Amer., vol. 85, 1989a, pp. 858-867.

Wightman, F. L.; and Kistler, D. J.: Headphone simulation of free-field listening II: psychophysical validation. J. Acoust. Soc. Amer., vol. 85, 1989b, pp. 868-878.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-103835		

4. Title and Subtitle	5. Report Date
Three-Dimensional Virtual Acoustic Displays	July 1991
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Elizabeth M. Wenzel	A-91061
	10. Work Unit No.
	505-67-01

9. Performing Organization Name and Address	11. Contract or Grant No.
Ames Research Center Moffett Field, CA 94035-1000	
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Technical Memorandum
National Aeronautics and Space Administration Washington, DC 20546-0001	14. Sponsoring Agency Code

15. Supplementary Notes

Point of Contact: Elizabeth M. Wenzel, Ames Research Center, MS 262-2, Moffett Field, CA 94035-1000, (415) 604-6290 or FTS 464-6290

16. Abstract

The development of an alternative medium for displaying information in complex human-machine interfaces is described. The three-dimensional virtual acoustic display is a means for accurately transfering information to a human operator using the auditory modality; it combines directional and semantic characteristics to form naturalistic representations of dynamic objects and events in remotely-sensed or simulated environments. Although the technology can stand alone, it is envisioned as a component of a larger multisensory environment and will no doubt find its greatest utility in that context. The general philosophy in the design of the display has been that the development of advanced computer interfaces should be driven first by an understanding of human perceptual requirements, and later by technological capabilities or constraints. In expanding on this view, the paper addresses current and potential uses of virtual acoustic displays, characterizes such displays, reviews recent approaches to their implementation and application, describes the research project at NASA Ames in some detail, and finally outlines some critical research issues for the future.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Virtual displays Localization Multimedia interfaces 3D sound	Unclassified-Unlimited Subject Category - 53

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	35	A03